

**This Page Is Inserted by IFW Operations
and is not a part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- **BLACK BORDERS**
- **TEXT CUT OFF AT TOP, BOTTOM OR SIDES**
- **FADED TEXT**
- **ILLEGIBLE TEXT**
- **SKEWED/SLANTED IMAGES**
- **COLORED PHOTOS**
- **BLACK OR VERY BLACK AND WHITE DARK PHOTOS**
- **GRAY SCALE DOCUMENTS**

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2003, 13:37:13 ; Search time 3200 Seconds

(without alignments) 10558.862 Million cell updates/sec

Title: US-10-054-562a-3

Perfect score: 1161

Sequence: 1 atgatgatctctctatctgc.....ttcgtccaatcagaagca 1161

Scoring table:

IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```
1: gb_ba:*
2: gb_hlg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_hlg_hum:*
31: em_hlg_inv:*
32: em_hlg_other:*
33: em_hlg_mus:*
34: em_hlg_pln:*
35: em_hlg_rod:*
36: em_hlg_mam:*
37: em_hlg_vrt:*
38: em_sy:*
39: em_hlgo_hum:*
40: em_hlgo_mus:*
41: em_hlgo_other:*
```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1161	100.0	1161	6	AR204520	Sequence
2	1161	100.0	1161	6	AR204521	Sequence
3	1161	100.0	1161	6	AR204518	Sequence
4	1161	100.0	1161	6	AR204519	Sequence
5	646.4	55.7	1380	3	AF125580	Wuchereria
6	504.2	43.4	1372	6	AR204522	Sequence
7	504.2	43.4	1372	6	AR204523	Sequence
8	461	39.7	892	6	AR204531	Sequence
9	461	39.7	892	6	AR204532	Sequence
10	339.2	29.2	25823	3	CEP2285	Sequence
11	323.8	27.9	34831	3	CEC47G2	Sequence
12	315	27.1	2310	3	CELCUT1	Sequence
13	310.4	26.7	813	6	AR204524	Sequence
14	310.4	26.7	813	6	AR204525	Sequence
15	249	21.4	5792	3	ALU73005	Sequence
16	217.2	18.7	2584	3	MAMTCUT1	Sequence
17	190.6	16.4	637	3	BMA012617	Sequence
18	156.6	13.5	358	3	BPA012618	Sequence
19	107.4	9.3	39478	3	CEP53F1	Sequence
20	85.6	7.4	56511	3	AC084467	Sequence
21	83.8	7.2	31536	3	CEP04D5	Sequence
22	74.8	6.4	32412	3	CEP53B6	Sequence
23	67	5.8	39736	3	CEP07E3	Sequence
24	62.6	5.4	36306	3	CEM442	Sequence
25	62.2	5.4	36089	3	CEZK265	Sequence
26	56.2	4.8	25284	3	CEW01A8	Sequence
27	55.4	4.8	59533	3	AC084450	Sequence
28	54.8	4.7	28396	3	U41264	Sequence
29	53.6	4.6	40265	3	CEM28	Sequence
30	52.8	4.5	34368	3	CEM06D12	Sequence
31	51.8	4.5	40600	3	CEP20D1	Sequence
32	49	4.2	26449	3	U23449	Sequence
33	49	4.2	31731	3	U21318	Sequence
34	48.4	4.2	327650	1	MPUM01	Sequence
35	48.2	4.2	53150	2	AC116984	Sequence
36	47	4.0	45713	3	AF067608	Sequence
37	46.8	4.0	27770	3	CEP22C8	Sequence
38	44.8	3.9	6942	1	AF455359	Sequence
39	44	3.8	229420	2	AL772373	Sequence
40	43	3.7	81091	9	AL645495	Sequence
41	43	3.7	164389	3	PFMAL3P6	Sequence
42	43	3.7	165318	2	AC126747	Sequence
43	42.8	3.7	48063	8	AF288090	Sequence
44	42.2	3.6	1141	6	AX083744	Sequence
45	41.4	3.6	79769	2	AC115678	Sequence

ALIGNMENTS

RESULT 1
LOCUS AR204520 1161 bp DNA linear PART 20-JUN-2002
DEFINITION Sequence 3 from patent US 6368600.
ACCESSION AR204520
VERSION AR204520.1 GI:21501867
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1161)
AUTHORS Chandrasekhar, R. and Morales, T. H.
TITLE Parasitic helminth cuticlin nucleic acid molecules and uses thereof
JOURNAL Patent: US 6368600-A 3 09-APR-2002;
FEATURES Location/Qualifiers

OY		481	GGACCAACCGGTCACAAGTTCCATTTGGTATCATGTGCAGCAGCTTATCATATAATGG	540
Dd		681	GGACCACACCGGTCAACCAATTGGTATCATGTGCAGCAGCTTATCATATAATGG	622
OY		541	ACATCGAATTCGAAGAACCCTTGTAATCTTCTCGCGGGTGTGCCAATCCCTTTGCGAT	600
Dd		621	ACATCGAATTCGAAGAACCCTTGTAATCTTCTCGCGGGTGTGCCAATCCCTTTGCGAT	562
OY		601	GATGGTAAACGGTGATACTGTGGAAATTCCTAAATGCTGATGATGTGCTTGTAATAAT	660
Dd		561	GATGGTAAACGGTGATACTGTGGAAATTCCTAAATGCTGATGATGTGCTTGTAATAAT	502
OY		661	TTCGTAATTAATTTGGGAATATCCAACAGATTTAATGGCTGGCCCAAGAGCTCACGATAC	720
Dd		501	TTCGTAATTAATTTGGGAATATCCAACAGATTTAATGGCTGGCCCAAGAGCTCACGATAC	442
OY		721	AATATGCGGAGATCAGATCAGAGCTTTTTCTATVCAAAGCCAGATCAGTATTAACATTAAGAA	780
Dd		441	AATATGCGGAGATCAGATCAGAGCTTTTTCTATVCAAAGCCAGATCAGTATTAACATTAAGAA	382
OY		781	CCAAATAGCGAATGTGTGTGACACAGAAGTTAGAAAACCAAGAGATGTGGACACTGTAA	840
Dd		381	CCAAATAGCGAATGTGTGTGACACAGAAGTTAGAAAACCAAGAGATGTGGACACTGTAA	322
OY		841	ACAGGTGTGTCGCGCAGCAAAACCTCTGTGAGCTGGCGCAACTCGTTTACTCAAGAAAAA	900
Dd		321	ACAGGTGTGTCGCGCAGCAAAACCTCTGTGAGCTGGCGCAACTCGTTTACTCAAGAAAAA	262
OY		901	TCTGAGAAACCGGAAATATCATGTATGTAGCAACTGATATCAACACCTTTAAATTAGC	960
Dd		261	TCTGAGAAACCGGAAATATCATGTATGTAGCAACTGATATCAACACCTTTAAATTAGC	202
OY		961	GATGATTAATCAAGCTTTTGCCAGTAGTTATTTAGCTACCGCGCACTCTGTGAACATTAATGA	1020
Dd		201	GATGATTAATCAAGCTTTTGCCAGTAGTTATTTAGCTACCGCGCACTCTGTGAACATTAATGA	142
OY		1021	CACCTGTAATACCTTGTCTGACGTACAAAATGSAATCTGCATGTACCAATTTGGCTTCCA	1080
Dd		141	CACCTGTAATACCTTGTCTGACGTACAAAATGSAATCTGCATGTACCAATTTGGCTTCCA	82
OY		1081	ATGTTATAGGGGTTTAAAGCATTCATTTGAATGTGTCGCGCATTTAATACATTGCGTTTAA	1140
Dd		81	ATGTTATAGGGGTTTAAAGCATTCATTTGAATGTGTCGCGCATTTAATACATTGCGTTTAA	22
OY	1141	TTTCGTCCAAATCAGAAGGCA	1161	
Dd		21	TTTCGTCCAAATCAGAAGGCA	1
<hr/>				
RESULT 3				
LOCUS	AR204518		1779 bp	DNA linear PAT 20-JUN-2002
DEFINITION	Sequence 1 from patent US 6368600.			
ACCESSION	AR204518			
VERSION	AR204518.1		GI:21501865	
KEYWORDS	.			
SOURCE	Unknown.			
ORGANISM	Unclassified.			
REFERENCE	1 (bases 1 to 1779)			
AUTHORS	Chandrasekar,R. and Morales,T.H.			
TITLE	Parasitic helminth cuticulin nucleic acid molecules and uses thereof			
JOURNAL	Patent: US 6368600-A; 1 09-APR-2002;			
FEATURES	location/Qualifiers			
source	1..1779 /organism="unknown"			
BASE COUNT	575 a 327 c 302 g 575 t			
ORIGIN				

OY	1	ATATGATTCGGCTTATTCGCTTCCTGTAACACCTTATTCGATTCGCTATTCGATTCGCG	60
Db	167	ATGATGATTCGGCTTATTCGCTTCCTGTAACACCTTATTCGATTCGCTATTCGATTCGCG	226
OY	61	GTTGCAATGCTGTCGAAAGTGAGCCAGAAATTTGAATGGACCACTTCAATTAACAAATC	120
Db	227	GTTGCAATGCTGTCGAAAGTGAGCCAGAAATTTGAATGGACCACTTCAATTAACAAATC	286
OY	121	AATTTTAATACAGTAATGCATTCGGAAGCAATGTTTATGTGAAGGTCCTTATATGATCAA	180
Db	287	AATTTTAATACAGTAATGCATTCGGAAGCAATGTTTATGTGAAGGTCCTTATATGATCAA	346
OY	181	GAAGGTTCGGCTATGATGAAAGTGAGCGTCAAGATTGCCGGAATTTCACTTCATTTGAT	240
Db	347	GAAGGTTCGGCTATGATGAAAGTGAGCGTCAAGATTGCCGGAATTTCACTTCATTTGAT	406
OY	241	TCATCAATGTTGGCGGACACGATCTCTGTAATTCACAGTGGTATTTTGTGAACAAACACT	300
Db	407	TCATCAATGTTGGCGGACACGATCTCTGTAATTCACAGTGGTATTTTGTGAACAAACACT	466
OY	301	GTTCGTAATTTGCTTTCATTCATTAATTTTGTACCAAAAGTTGATCGTCATATGAGTACAA	360
Db	467	GTTCGTAATTTGCTTTCATTCATTAATTTTGTACCAAAAGTTGATCGTCATATGAGTACAA	526
OY	361	TGCTTTTACATGGAAGCTGATAAAACAGTTAGTCACACATTTAGSTATCTGAAATCACA	420
Db	527	TGCTTTTACATGGAAGCTGATAAAACAGTTAGTCACACAGTATGAGSTATCTGAAATCACA	586
OY	421	ACTGCTTTTCAAACTCAAAATTTGCCGATCCGATATGCGCTTATGTAATTTTGGATGCT	480
Db	587	ACTGCTTTTCAAACTCAAAATTTGCCGATCCGATATGCGCTTATGTAATTTTGGATGCT	646
OY	481	GGACCAACCGGTCAACACAGTTTCATTTGCTATCATTTGGTCAGCCAGTTATCATATAATG	540
Db	647	GGACCAACCGGTCAACACAGTTTCATTTGCTATCATTTGGTCAGCCAGTTATCATATAATG	706
OY	541	ACATGCGATTTCTGAACCCGTTGATCTTCTGCGCGGATGTCCATTCCTGCTTGTGAT	600
Db	707	ACATGCGATTTCTGAACCCGTTGATCTTCTGCGCGGATGTCCATTCCTGCTTGTGAT	766
OY	601	GATGGTAAACGGTGTACTGTGGGAAATTTCTAAAGCTGATGAGATGCTCTTGATATAAT	660
Db	767	GATGGTAAACGGTGTACTGTGGGAAATTTCTAAAGCTGATGAGATGCTCTTGATATAAT	826
OY	661	TTTGCTAAATATTTTGAATATCCACAGATTTATGGCTGGCCAAACACTCAGATATC	720
Db	827	TTTGCTAAATATTTTGAATATCCACAGATTTATGGCTGGCCAAACACTCAGATATC	886
OY	721	AAATATGCGGATCGATCAGCTTTCTCTATCAATGCCAGATCAGTATTAACATTAAGAA	780
Db	887	AAATATGCGGATCGATCAGCTTTCTCTATCAATGCCAGATCAGTATTAACATTAAGAA	946
OY	781	CCAATATGCGGAATGTGTGCACCAATATGTTGGAACCAACAGAGATTCGGAGCTGTAA	840
Db	947	CCAATATGCGGAATGTGTGCACCAATATGTTGGAACCAACAGAGATTCGGAGCTGTAA	1006
OY	841	ACAGAGTGGTCCGAGCAAAACCTGCTGCAGCTGCCCACTTCGTTATCTCAAGAAAAA	900
Db	1007	ACAGAGTGGTCCGAGCAAAACCTGCTGCAGCTGCCCACTTCGTTATCTCAAGAAAAA	1066
OY	901	TTCTGACAGACCGGAGAATATCATTTGATGTAGCAAGATATCAACACCCCTGAAATATG	960
Db	1067	TTCTGACAGACCGGAGAATATCATTTGATGTAGCAAGATATCAACACCCCTGAAATATG	1126
OY	961	GATATATATCAAGCTTTGCCAGTTGATTTACAGTACCGGCACTTCGTGCAATATATGGA	1020
Db	1127	GATATATATCAAGCTTTGCCAGTTGATTTACAGTACCGGCACTTCGTGCAATATATGGA	1186
OY	1021	CAACCTTAAATACCTTGCATGACATGACAAATTTGAATTCGATCTCACATTTGGCTTCA	1080
Db	1187	CAACCTTAAATACCTTGCATGACATGACAAATTTGAATTCGATCTCACATTTGGCTTCA	1246
OY	1081	ATGTTTATGGGCTTAAACATTTGCATTTGATTCCTGCCGTCATTAATTAACATTTGTTTAAA	1140

```

Db 1247 ATGTTATAGGTTAAAGCATGATGATGCGCGCATTTATTAACATTTGTTTAA 1306
Oy 1141 TTTCGTCAAAATCAGAAGCA 1161
Db 1307 TTTCGTCAAAATCAGAAGCA 1327

RESULT 4
AR204519/c 1779 bp DNA linear PAT 20-JUN-2002
LOCUS AR204519 Sequence 2 from patent US 6368600.
DEFINITION AR204519
ACCESSION AR204519
VERSION AR204519.1 GI:21501866
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1779)
AUTHORS Chandrasekar, R. and Morales, T. H.
TITLE Parasitic helminth cuticulin nucleic acid molecules and uses thereof
JOURNAL Patent: US 6368600-A 2 09-APR-2002;
FEATURES
source 1..1779
location/Qualifiers
BASE COUNT 575 a 502 c 327 g 575 t
ORIGIN

Query Match 100.0%; Score 1161; DB 6; Length 1779;
Best Local Similarity 100.0%; Pred. No. 6,3e-262;
Matches 1161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 1ANGATGATTCGTTCTTATTCCTTTCGTACTACATTTATTCATTTGTTTCGATTCG 60
Db 1613 1ANGATGATTCGTTCTTATTCCTTTCGTACTACATTTATTCATTTGTTTCGATTCG 1554
Oy 61 6TTGCAATGGTGCAGAGTGCAGCAAAATGATGAGCAACCTTCATTAAC 120
Db 1553 6TTGCAATGGTGCAGAGTGCAGCAAAATGATGAGCAACCTTCATTAAC 1494
Oy 121 1AATTTAATACAGTAAATCATTCGAAGACATGTTTATGTAAGGCTTTATGATCAA 180
Db 1493 1AATTTAATACAGTAAATCATTCGAAGACATGTTTATGTAAGGCTTTATGATCAA 1434
Oy 181 6AAGGTTCGCGTAATGATGAGGTGAGCTCAAGTTGCCGAATTTCACTTCATTTGAT 240
Db 1433 6AAGGTTCGCGTAATGATGAGGTGAGCTCAAGTTGCCGAATTTCACTTCATTTGAT 1374
Oy 241 1CATGCAATGTCGCGGTACAGATCTCGAATCCACGATGATTTTGTAAACAACACT 300
Db 1373 1CATGCAATGTCGCGGTACAGATCTCGAATCCACGATGATTTTGTAAACAACACT 1314
Oy 301 6TTTGCATTTCTTTTATTCATTTATTTGTACCAAAAGTTGATCGATGAGTACAA 360
Db 1313 6TTTGCATTTCTTTTATTCATTTATTTGTACCAAAAGTTGATCGATGAGTACAA 1254
Oy 361 1GCTTTTACATGGAAGCTGATAAACAGTTAGTGCAAGATGAGGTATCTGAATACAA 420
Db 1253 1GCTTTTACATGGAAGCTGATAAACAGTTAGTGCAAGATGAGGTATCTGAATACAA 1194
Oy 421 1ACTGTTTTCACAACTCAATTTGTCGCCGATGAGTATGCGGTTGTAAGAAATTTGGATGT 480
Db 1193 1ACTGTTTTCACAACTCAATTTGTCGCCGATGAGTATGCGGTTGTAAGAAATTTGGATGT 1134
Oy 481 6GACCAACCGGTCAACAGTTCAATTTGCTATCATTTGTCAGCCAGTTATATCAATAATG 540
Db 1133 6GACCAACCGGTCAACAGTTCAATTTGCTATCATTTGTCAGCCAGTTATATCAATAATG 1074
Oy 541 1ACATGCGATTTGAAACCGTTGATCTTTCGCGGTTGTCATTCCTGCTTGTTCGAT 600
Db 1073 1ACATGCGATTTGAAACCGTTGATCTTTCGCGGTTGTCATTCCTGCTTGTTCGAT 1014
Oy 601 6ATGTTAAGCGTATCTGTGAAATTTCTAAATGCTGATGATGTGCTCTTGAATAATAT 660

```

```

Db 1013 GATGTAAGCGTATCTGTGAAATTTCAAAGCTGATGATGCTCTGTTAATATAT 954
Oy 661 1TTGCTAATTAATTTGAATATCCACAGATTTAATGGTGGCCCAAGATCAGTATAC 720
Db 953 1TTGCTAATTAATTTGAATATCCACAGATTTAATGGTGGCCCAAGATCAGTATAC 894
Oy 721 1AATATGCGGATCGATCAGAGCTTTTCTATCAATGCGAGATCAGTATTAACATTAAGA 780
Db 893 1AATATGCGGATCGATCAGAGCTTTTCTATCAATGCGAGATCAGTATTAACATTAAGA 834
Oy 781 1CCAAATAGGAATGTTTGTGACCAACATGTTGAAACCAACAGATTTGCGAGCTGTAA 840
Db 833 1CCAAATAGGAATGTTTGTGACCAACATGTTGAAACCAACAGATTTGCGAGCTGTAA 774
Oy 841 1ACAGTGTGCGCCAGCAAAACCTGCTGAGCTGCGCAACTTCGTTTACTCAAGAAAGA 900
Db 773 1ACAGTGTGCGCCAGCAAAACCTGCTGAGCTGCGCAACTTCGTTTACTCAAGAAAGA 714
Oy 901 1TTCGCAAAACCGAGAAATATCATTTGATGATGATGATGATGATGATGATGATGATG 960
Db 713 1TTCGCAAAACCGAGAAATATCATTTGATGATGATGATGATGATGATGATGATGATG 654
Oy 961 1GATGATTAATCAAGCTTTGCGAGTGTGATTTACGTACACCGTGCACCTTGTGCAATATGA 1020
Db 653 1GATGATTAATCAAGCTTTGCGAGTGTGATTTACGTACACCGTGCACCTTGTGCAATATGA 594
Oy 1021 1CAACGTATATACCTGCTGAGTAAATGCAAAATGATGATGATGATGATGATGATGATG 1080
Db 593 1CAACGTATATACCTGCTGAGTAAATGCAAAATGATGATGATGATGATGATGATGATG 534
Oy 1081 1ATGTTATGGGTTTAAAGATTCATTTGATGATGATGATGATGATGATGATGATGATG 1140
Db 533 1ATGTTATGGGTTTAAAGATTCATTTGATGATGATGATGATGATGATGATGATGATG 474
Oy 1141 1TTTCGTCAAAATCAGAAGCA 1161
Db 473 1TTTCGTCAAAATCAGAAGCA 453

RESULT 5
AF125580 1380 bp mRNA linear INV 01-AUG-2000
LOCUS AF125580 Wuchereria bancrofti cuticulin-1 (cut-1) mRNA, complete cds.
DEFINITION AF125580
ACCESSION AF125580
VERSION AF125580.2 GI:9625063
KEYWORDS
SOURCE Wuchereria bancrofti.
ORGANISM Wuchereria bancrofti.
REFERENCE 1 (bases 1 to 1380)
AUTHORS Ramzy, R., Helmy, H., Adely, M., Curtis, K. and Well, G.
TITLE Wuchereria bancrofti L3 cuticulin-1 cDNA
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1380)
AUTHORS Ramzy, R., Helmy, H., Adely, M., Curtis, K. and Well, G.
TITLE Direct Submission
JOURNAL Submitted (03-FEB-1999) Research & Training Center on Vectors of Diseases, Ain Shams University, Abassia Square, Cairo, Egypt
REFERENCE 3 (bases 1 to 1380)
AUTHORS Curtis, K., Ramzy, R., Helmy, H., Adely, M. and Well, G.
TITLE Direct Submission
JOURNAL Submitted (15-JUL-1999) Infectious Diseases, Washington University School of Medicine, 216 S. Kings Highway, St. Louis, MO 63110, USA
REMARK Sequence update by submitter
COMMENT On Aug 1, 2000 this sequence version replaced gi:4741874.
FEATURES
source 1..1380
location/Qualifiers
/organism="Wuchereria bancrofti"
/db_xref="taxon:6293"
/dev_stage="L3 larvae"
/country="Egypt; Nile Delta"

```

gene	1..1380	
/gene="cut-1"	108..1280	
CDS	/gene="cut-1"	
	/note="similar to Ascaris lumbricoides cuticulin-1"	
	/codon_start=1	
	/product="cuticulin-1"	
	/protein_id="AAD28743.2"	
	/db_xref="GI:9625064"	
	/translaton="MLRISILIVAFITIVAPVDNVEGEPEIEGCTSTTYNTR	
	NPEFGHYVVGSLYDQGCNRNDEGGRQVAGIELPDDSCNVARHSSINRGIFVVTYNT	
	PSGPLEFATIKDRAVRVOCQFYMENADKPTVRSIVSEITTAATQIVPAPVGRYIELDGG	
	PSGPGPQFATIGQPVYVYHKTKCDSSTVTPFAVYVHSCVDGNGKVELLNADCAIDK	
	FLNLNLEIPTDIDMAGQZAHYIKIADRSQLTFOCCITITIEPNSECRPOCTEPQGG	
	AIKTRNGVAALALROAAAFRVYLRKRDVDENIVDVRIDLVNADINEPTSLPNLRHS	
	SLAHNGEHVIVATVMQGLICMSTGTGTAGMLIFVYSVAITVAITVALILRSHSTKV"	
BASE COUNT	456 a 224 c 244 g 454 t 2 others	
ORIGIN		

Query Match	55-78;	Score	646.4;	DB	3	Length	1380;
Best Local Similarity	74.48;	Pred.	No. 3, 8e-141;				
Matches	844;	Conservative	0;	Mismatches	281;	Indels	9;
						Gaps	2;

QY 1 ATGATGATTTCGGTCTTATGCTTTCTGTACTACCTTATAGCATGTCTTATTCGATCCG 60
|| || | || | || || || || || || || || || || || || || || || || || || || ||
Db 105 ATTATGCTTCGATTATATAGCATATTAGTAGCACTTTATTACTATTGTAAATAGCATCCG 164

0y GTTGACATGGTGTGCGAAGGTGCAGCCGAATAATGAATGTGGACCACCTTCATATACAAATC 120
|||||
61
Db GTTCATATATGGTGTGCGAAGGTCAACCGCAATATTGAATGTGTGTCACCAATCATTAACGGTT 224

09 ATTTTATACAGCTATGCATTCGAAGACATGTTATGTGAAGGCTTATGATCAA 180
 |||||
 121 ATTTTATACAGCTATGCATTCGAAGACATGTTATGTGAAGGCTTATGATCAA 180
 |||||
 225 AATTTTACACCTCGAAATCCATTTTGAAGACATGTTATGTAAAAGGATTTATGATCAA 284
 |||||

QY	181	285	Db
GAAGTGGCCGGATATGATGAAGTGGACGCTCAAGTTGGCCGAATTTCACTTCCATTGAT	240		
CAAGGATGTCGATATGATGAAAGTGGACGACAGTAAAGTGGAAATTTGAGCTACCACTTTGAT	344		

9y	241	TCAGCAGTGTGGCGGTACAGATCTCTGAATCCAGCTGTATTTTGTGAACACAACT	300
Db	345	TCATGTATGTACACATCACCTCTTGTTAAATCCAGCTGTATATTTTGTCCAAACACA	404

	301	405
Oy	GTTCATATTCGCTTCATCCATATTTCTTACCAAGTTGATCGCATATCCAGTACAA	360
Db	GTGTCATTTCTTCATTCATATTTGTGCACAAAGTTGATCGCATATTCGTGTACAA	464

Qy	Db
361	465
TGCTTTACAGGAAGCTGATTAACAGTTAGTCACACAGATTGAGGTATCTGAATCACA	TGCTTTACATGGAAAGCTGATTAACACTGTTACACACACAAATTTGAAGTTTCCGAAATATCA
420	524

Qy 421 ACTGCTTTTCAAACTCAAATTTGTCCCGATGCCAGTATGCCGTTATGAAATTTTGGATGTT 480
Db 525 AAGCAATTTGCACACACAAATTTGTATCCCAATGCTGTATGTGCATATGAGATCTTTGATGTT 584

QY	481	GGACCAACCGGTCACCAAGTTCATTTGGTCATCATTCGTCAGCCAGTTTATCATAAATGG	540
Db	585	GGACCATTCAGGACCAACCGGTTTCATTTGTCACCATTCAGGACCAACCGGTTTATCATAAATGG	644

QY	541	ACATGCGATTTCGAAACCGTTGATCTTCTGCGCGGTTGTCATTCCCTGCTTTGTCGAT	600
Db	645	ACATGCGATTTCGAAACCGTTGATCTTCTGCGCGGTTGTCATTCCCTGCTTTGTCGAT	704

QY	601	GATGTAACGGTGATACTGTGGAAATTCTAATAGTCGATGATGTGCTTGTGATAAATAT	660
Db	705	GATGTAATGGTGTGAATGAAGTAGAGCTATTAAATGCATGATGTTGTGCTTTGGACAATTT	764

QY 661 TTGCTAAATTAATTGGAAATATCCACAGATTTAATGCTGGCCCAAGAGCTACGTATAC 720
765 CTACCTTAATTAATTTAGAAATATATCCACAGATTTAATGCTGGCCCAAGAGCACACCGTTAC 824

721 AAATATGGGATCGATCACAGCTTTCTATCATCGACGATCGATATACCATTAAGAA 780

Db	825	AAATACGCTGATCGATTCACAACTTTTTTATCAATGTCAATATAGCATTAACATTAATTAAGAA	884
QY	781	CCAATAGCGAATGTGTTTGACCCACAATGTTAGAACCAANGATTCGGAGCTGTTAAA	840
Db	885	CCAAATATGTGATGTCAACGACCAACAATATGCATGAACCCCAAGGTTTGGCCCAATTAAA	944
QY	841	ACAGG---TGGTGCAGCAGCAAAACCTGCTGAGCGCCCACTTGCTGTATTCAGAA	897
Db	945	ACGCCAATGGTGTGTTGCACGTGCTCATTAGCAGCAAGCTGCTGCATTTAGATATTGAAAAAG	1004
QY	898	AGATTCGACAGAACCGAGAAATATCATGTGATGTAGCAACTGATATCAACACCCCTGAAAT	957
Db	1005	AGAGATGTAGAGATGAAATATTTGTGATGTACCTACCGATCTTAATCATTTGATGATATA	1064
QY	958	AGCGATGATTAATCAAGCTTTTGCCAGTTGATTTAGCTACGCCCGTGCACTTCG-----CAA	1011
Db	1065	AATGAAGATTTACTCTCATTTACCAAAATGGCGCTAGCGTCAATCATTTGTTAGCACAT	1124
QY	1012	CATAATGGACAACTGTATATCTGCTCCAGTACAAAAATGSAATCTGCATGTCCACATTT	1071
Db	1125	GAAATAGTCAATCCGGTATATGTTTCAACAATGACACAAGATATATGATATTCATTAAC	1184
QY	1072	GGCATTCGAATGTTATGGGTTTAAGCATGTGATGATTGCTGCCGTATTATT	1125
Db	1185	GGTTTACACTGCGCGATAGCTAATATTTGTTATGTTTCTGTGCTACAAATT	1238

RESULT 6	
AR204522	
LOCUS	
AR204522	
1372 bp	
DNA	linear
PAT 20-JUN-2002	

DEFINITION Sequence 6 from patent US 6368600.
 ACCESSION AR204522
 VERSION AR204522.1 GI:21501870
 KEYWORDS .

SOURCE	ORGANISM
Unknown.	Unknown.
Unknown.	Unclassified.
REFERENCE	1 (bases 1 to 1372)

AUTHORS Chandrasekhar, R. and Morales, T.H.
 TITLE Parasitic helminth cuticular nucleic acid molecules and uses thereof
 JOURNAL Patent: US 6368600-A 6 09-APR-2002;
 REATUES Location/Qualifiers

source	1. .1372	/organism="unknown"
BASE COUNT	456 a	229 c
ORIGIN	281 g	406 t

Query Match	43.4%	Score 504.2	DB 6	Length 1372
Best Local Similarity	76.9%	Pred. No. 9e-108		
Matches 652: Conservative	0	Mismatches 193	Indels 3	Gaps 3

[illegible]

60 GGTGACATAGTGTGTCGAAGCTGACCCAGAAATGATGTGACCACCTTCATAACAT 119
 |||||
 92 TATTGACCAATGCTGTGGAAGCTGAACTGTAATAGATGTGTGGCCGCTTCGTAACAT 151
 |||||

OY 120 CAATTTAATAACACCGCAATGTCATTCGAAGACATGTTATGTGAAGGCTTATGATCA 179
|||||
152 CAATTCAATTAAGTAGAAATCATTTTGAAAGCACAGTATATTAAGAAGCACTTATATCA 231
|||||

180 AGAAGCTTCCCGCTAATGATGATGAAGGTGGACGTCACATGGCCGAATTTCACCTCCATTTGA 239
 212 GGATGATATGCGCTGCTTCGATAGTAATTCAGACCGGTACAGTAACTGCTGAAATTCGCAATTCGG 271

240 TTCAATGCAATGTTGGCGGTACACAGATCTCTGATCCACGTCGTATTTTGTAAACAAAC 299
 272 TTCTGCTAAATTTTGAACGATTCACGATTCCTTTAAATCTCCGCTGCTCTTTTCTTCTCTCC 331

300 TGTGTCAATTGGTTTCATTCATTTTGTGTACCAAGTGCATCGCATTCGAGTACA 359
||||||| | ||||| || ||||| ||| ||||| ||||| |||||

DB	332	ACTTGTCATTTCAATTTCCAAAAATTTGGTTACAAAATAATGATGCAGATATCGTATACA	391
QY	360	ATGCTTTTACATGC -AAGCTGATTAACAGATTAGTGC -ACAGATTGAGGTATCGAAATC	417
DB	392	AGTTTATATATGAAGAGCTGATTAAGACCGTATAGTACGTCTTCCGAAATCTGTAAATG	451
QY	418	ACAATGCTTTTCAACACTCAAAATTTGCCGATGCCAGTATGCCGTTATGAAATTTGGAT	477
DB	452	ACTACAGCATTTCCAAACTCAAGTGGTATACCAATGCCCGTATGTGCATATAGATATTGGAA	511
QY	478	GGTGACCAACCGGTCAACCACTTCATTTGCTATCATTTGGTTCAGCAGTTTATCATAA	537
DB	512	GGTGACCAACCTGGTGCACCTCTGTGATTTGCCAAATGATGGAGATCATGTATACAAA	571
QY	538	TGGAGATGGATTTTCGAAAACCGTATATCTTTCTGCGGGGTGTCCATCTCGCTTGTTC	597
DB	572	TGGACATGTGATTCAGACAGTACAGATCATTTCTGTGCATTATGATCATCATGTGTGTG	631
QY	598	GATGATGGTAAACGATGATCTGTGAAAATTCCTAAATGCTGATGATGTGCTTTGATAA	657
DB	632	GATGATGGAAAAGTGATGACAGTGGAGATTTTGATGATGAAGAAGATGTGCTTTGACAA	691
QY	658	TATTTTGCTTAATATTTTGGATATTCACACAGATTATATGCGTGGCCAGAAAGCTACGTA	717
DB	692	TATTTTACCTCAAAATTTTGGAAATATTTACAGATTTAAATGGCGGCAAGAGCTATGTT	751
QY	718	TACAATATGCGGATGCAGTACACAGCTTTTCTCATGATCCAGATCAGTATTCATTAAT	777
DB	752	TATTAATATGCGAGATGCAGATCAGATCTTACATACATGCTCAGATATGATTAATCAAT	811
QY	778	GAAACCAATAGCGAATGTGTTGACCCACAAATGTTCAAGACCAACAAAGATTGGAGCTGT	837
DB	812	GAGCCACATATGAGGATGTCTCGACCAATATGCACAGAGCCACAAAGGATTTGTGCCATA	871
QY	838	AAAACAG 845	
DB	872	AAATCTGG 879	
RESULT 7			
LOCUS	AR204523	1372 bp	DNA linear PAT 20-JUN-2002
DEFINITION	Sequence 7 from patent US 6366600.		
ACCESSION	AR204523		
VERSION	AR204523.1	GI:21501871	
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	1 (bases 1 to 1372)		
AUTHORS	Chandrasekhar,R. and Morales,T.H.		
TITLE	Parasitc helminth cuticlin nucleic acid molecules and uses thereof		
JOURNAL	Patent: US 6368600-A 7 09-APR-2002;		
FEATURES	Location/Qualifiers		
source	1..1372		
BASE COUNT	406 a 281 c 229 g 456 t		
ORIGIN			
Query Match	43.4%; Score 504.2; DB 6; Length 1372;		
Best Local Similarity	76.9%; Pred. No. 9e-108;		
Matches	652; Conservative 0; Mismatches 193; Indels 3; Gaps 3;		
QY	1	ATGATGATTCGTCCTTATTCGCTTCTGTACTACACTTATTG-CATTGCTTATTTGCATTCC	59
DB	1341	ATTATCATTTATTCCTCTCTGTTATTCCTACTTCTTACTTGAAGTATACATATCATGATCC	1282
QY	60	GATTGACATGATGTCGAGAGGACCGAAGTAATGATGAGCAACTTCATTAACAT	119
DB	1281	TATTACATATGTCGAGAGGACCTGAAATAGAAATGTGGCGGACGTTTGATATCAAT	1222
QY	120	CATTTTAAATACAGCTAAATGCAATTCGAAGACATGTTATGTAAGAGTCTTTATGATCA	179

Db	1221	CAATTTCATACCTAGAAAATACATTTTGAAGGACACGATATATGTAAAAGAGCTGTATGATCA	1162
QY	180	AGAGAGTTCGCCCTAATGATGAAAGGTGGACGCTAAAGTTGCCGGAATTTTCACTTCCATTGGA	239
Db	1161	GGATGATATCTCGTTACAGATAGTAATGAGCGGACGAGTAGCTGGAATTCGAAATGGCAATGGA	1102
QY	240	TTTCATTCATGTTGGCGCGTACAGATCTCTGTATCGATCCAGCTGGTATTTTGTAAACAAC	299
Db	1101	TTTCGTTAATGTTTGAACGATTCACGATCCTTAATCTCTGSGTGTTTTTTGTAAACAACGT	1042
QY	300	TGTTGTCAATTCGTTTCATCCATTTATTTGTACCAAAAGTTGATGCTGATATCGAGTACA	359
Db	1041	AGTTTGCAATTCATTTTCATTCACAAAATTCGTTTCCAAAAATAGATCCAGAGATATCGATAC	982
QY	360	ATGCTTTTTCATAGG-AAGCTGATTAACAGCTAGAGG-ACAGATTTGAGGTATCTGAATTC	417
Db	981	ATGTTTATTTATGGAAGAGGTGATTAAGACCGCTTAGTACTGTGCTTGAAGTATCTGAATG	922
QY	418	ACAACTGCTTTTCAAACTCAAAATTTGCTCCGATGCCAGATATGCCGTTATGAAATTTTGGAT	477
Db	921	ACTGACGATTCCTCAAACTCAAGTGGTACCAATGCCCGATGTGCGATATGAGATTTTGGAA	862
QY	478	GGTGGACCAACCGGTACACCACTTCAATTTGCTATCATTTGGTCCAGCCAGTTATCATATAA	537
Db	861	GGTGGACCAACCGGTGGACACCTTGTGATTTGCAATGATTTGGAGATCATGTATATACAAA	802
QY	538	TGGACATCGGATTTTCGAAACCGTTATATCTTCTCGCGGGTGTGCTCATCTCGCTGTCTC	597
Db	801	TGGACATGTGATTTACAGAGACTACAGATACATCTCTGTGATTTAGTACATTCATCATGTGTGG	742
QY	598	GATGATGTTACCGGTGATATCTGTGGAAATTCGTAATGCTGATGATGTGTCTTGTATATAA	657
Db	741	GATGATGTTGAAAAAGGTGATGTGACGTGGAGATTCGAAATGAAAGGATGTGCTTTGGACAAA	682
QY	658	TATTTGCTTAATTAATTTTGGATATATCCAAAGATTAATTAAGGCTGGGCCAAGAACCTACGTA	717
Db	681	TTTTTACTTAATTAATTTTGGATATATTTACAGATTTAATTAAGCTGGGCCAAGAACCTATGTT	622
QY	718	TACAAATATGCGATCGATCACAGCTTTTCTATCAATGGCCAGATCAGTATTAACATTTAAA	777
Db	621	TATTAATATGACAGATCGATCACAGACTTTTCTATCAATGGCCAGATCAGTATTAACATTTAAA	562
QY	778	GAAACCAATAGCGATGTGTGACACCAATGTTTACAGAACCAACAGAGTTCGAGACTGTT	837
Db	561	GAGCCACATAGGGAATGTCTCGACCAATGCACAGAGCCACAGAGATTTGGTGTCCATTA	502
QY	838	AAAACAGG 845	
Db	501	AAATCTGG 494	
RESULT 8	AR204531		
LOCUS	AR204531	892 bp	DNA
DEFINITION	Sequence 16 from patent US 6368600.	linear	PAT 20-JUN-2002
ACCESSION	AR204531		
VERSION	AR204531.1	GI:21501881	
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	1 (bases 1 to 892)		
AUTHORS	Chandrashekar, R. and Morales, J. H.		
TITLE	Parasitic helminth cuticulin nucleic acid molecules and uses thereof		
JOURNAL	Patent: US 6368600-A 16 09-APR-2002;		
FEATURES	Location/Qualifiers		
source	1..892		
BASE COUNT	290 a 142 c 168 g 292 t		
ORIGIN			
Query Match	39.7%; Score 461; DB 6; Length 892;		
Best Local Similarity	78.1%; Pred.No. 1.3e-97;		

Query Match	39.7%;	Score 461;	DB 6;	Length 892;
Best Local Similarity	78.1%;	Pred. No. 1.3e-97;		

RESULT 10	CEP22B5/C	25823 bp	DNA	linear	INV 23-JAN-2002
LOCUS	CEP22B5				
DEFINITION	Caenorhabditis elegans cosmid F22B5, complete sequence.				
ACCESSION	Z50044				
VERSION	Z50044.1	GI:899234			
KEYWORDS	HTG; Cuticulin; Elongation factor; GMP-binding ADP-ribosylation factor; Phenylalanyl-tRNA synthetase; RNA binding.				
SOURCE	Caenorhabditis elegans.				
ORGANISM	Caenorhabditis elegans.				
REFERENCE	1				
AUTHORS	none.				
TITLE	Genome sequence of the nematode <i>C. elegans</i> : a platform for investigating biology. The <i>C. elegans</i> Sequencing Consortium				
JOURNAL	Science 282 (5396), 2012-2018 (1998)				
MEDLINE	99069613				
PUBMED	9851916				
REMARK	The <i>C. elegans</i> Sequencing Consortium.				
REFERENCE	2 (bases 1 to 25823)				
AUTHORS	Sims, M.A.				

TITLE	Direct Submission
JOURNAL	Submitted (13-JUL-1995) Nematode Sequencing Project, Sanger Institute, Hinxton, Cambridge CB10 1SA, England and Department of Genetics, Washington University, St. Louis, MO 63110, USA. E-mail: jes@sanger.ac.uk or rvenemacode.wustl.edu
COMMENT	Coding sequences below are predicted from computer analysis, using predictions from GeneFinder (P. Green, U. Washington), and other available information. Current sequence finishing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note. IMPORTANT: This sequence is not the entire insert of clone F22B5. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions. The true left end of clone F22B5 is at 24607 in sequence Z66522. The true right end of clone F22B5 is at 16002 in sequence Z66523. The true left end of clone M0506 is at 25723 in this sequence. The true right end of clone F14B5 is at 4610 in this sequence. The start of this sequence (1..101) overlaps with the end of sequence Z66522. The end of this sequence (25723..25823) overlaps with the start of sequence Z66523. For a graphical representation of this sequence and its analysis see: http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=F22B5 IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions. Location/Qualifiers
source	1..25823 /organism="Caenorhabditis elegans" /db_xref="taxon:6239" /chromosome="11" /clone="F22B5" join(411..462,584..663,745..1127,1216..1471) /gene="F22B5.2" join(411..462,584..663,745..1127,1216..1471) /gene="F22B5.2" /codon_start=1 /protein_id="CAA90354.1" /db_xref="GI:3876226" /db_xref="SWISS-PROT:Q19706" /translation="MAPAEVVSMAEAEQDNAPHIQGADGTRTETAFTEVDGVNRK VVQKRVINKRPKVVADRRKKVKFGSGEPQVATTVAAEEVDQAFRRNGEG IILDVEDKOTAKTTSREHRCRHKGDHNSCTGPKVMYKQIDLEADADKQTPRRAG MRPDGROIDRNSDENTCRVNLPOEMNEDELDLDFGKIGRIVIRFIARDVVTGLPKG FAFVFEESRDAAARAIAELNDIRMYHMYLKEWTRPSN"
gene	join(1870..2158,2209..2339,2534..2668) /gene="F22B5.1" join(1870..2158,2209..2339,2534..2668) /note="contains similarity to Pfam domain: PF00025 (ADP-ribosylation factor family), Score=295.0, E-value=2.9e-85, N=1" /codon_start=1 /protein_id="CAA90353.2" /db_xref="GI:14530408" /db_xref="SWISS-PROT:Q19705" /translation="WGFLKILRKORARERERILLIGDNGAKTTLMKKFLDEPTDITI EPTIGFDIKYVHFQDLNMDVGGQKLSRYWKMYFESTDALIMVYDSSDREKLQD
CDS	
gene	SEELKLLGEERLAGASLLVLANKSDLPGAIDVNSIAQVLDLHSIKSHMKIFSCAL SGRILQOATMTWICDDVGSRLFTLD" complement(join(2833..3042,3088..3700,3823..4048,4273..4393)) /gene="F22B5.3" complement(join(2833..3042,3088..3700,3823..4048,4273..4393)) /note="similar to cuticulin CDNA EST EMBL:701970 comes from this gene CDNA EST CEESX90RB comes from this gene CDNA EST EMBL:700412 comes from this gene CDNA EST yk386c7.3 comes from this gene CDNA EST yk386c7.5 comes from this gene CDNA EST yk504b12.3 comes from this gene CDNA EST yk543g2.3 comes from this gene CDNA EST yk552b8.3 comes from this gene CDNA EST yk504b12.5 comes from this gene CDNA EST yk543g2.5 comes from this gene CDNA EST yk551h11.5 comes from this gene CDNA EST yk590c10.5 comes from this gene CDNA EST yk652b8.5 comes from this gene" /codon_start=1 /protein_id="CAA90355.1" /db_xref="GI:3876227" /db_xref="SPTREMBL:Q19707" /translation="MARSLGIGLGLDLVAVSAIPVNNVGEPEVECGPTSTVYNN TRNATGEVTVKGLFDQOECRDEBGRVAGIETLPTFCVARTKRLNPKGVFVTTV VVSFHPQVTVKADRVAVQCFEMADKTVSQIIEVSDLTTFAPQVPMPICKCEILN GGTPDEPOVAFATIGQVYVHKWTCSEYDFECVAVHSCVDGNDVYQIIDEACAL DKFLNLANEYPTDLMAAGOEAAHYKAYADRSOLFQCOISITVKEENEECARPTCEPOG FGAVYQANOTPOPFVRLKRSAPVMEIQLDVAELTTEVEGLNPPSITQAOALVRS REIGDSRQELCISSEFHSITVYVTEICTVYVAFITIMYISRMVPSDKMQSAC"
gene	join(9651..9880,9940..10234) /gene="F22B5.4" join(9691..9890,9940..10234) /note="CDNA EST yk304c8.3 comes from this gene CDNA EST yk304c8.5 comes from this gene CDNA EST yk594h3.3 comes from this gene CDNA EST yk594h3.5 comes from this gene" /codon_start=1 /protein_id="CAA90356.1" /db_xref="GI:3876228" /db_xref="SPTREMBL:Q19708" /translation="WLSLKARRTKMPRVLSRNASTSVSPARIQLKKPAEAGHPT SRNNSRDRFVKVAVIQLNGDTPYQFLVRLGHAIEVYPLFVLTAAEVLFCASASTWSFG KAEIWLDRSNSKAPWDERLRLDYWKMPYVAFDLDGRTKRCELMEDLODEMLEAAK RQTR"
CDS	
gene	complement(join(10849..10984,11033..11343,11397..11521,11578..11832,11917..12154)) /gene="F22B5.5" complement(join(10849..10984,11033..11343,11397..11521,11578..11832,11917..12154)) /codon_start=1 /protein_id="CAA90357.1" /db_xref="GI:3876229" /db_xref="SPTREMBL:Q19709" /translation="MNCCKQASKAVKFLVGGFNKASRCFSEKLIIPKDLIKYKNPTS EQIKFQKLVGDVNDIKLKEETMKNDYQLVLCRKAKWSJALACYNINLNSNYSIY SDSVLSGTOSITYKSLNRSIPDQTFGLSYQPDNTYNLPHLHMEATDFCSVKMS GGCVDYSVAAMVRKVLTKVRGPTYYVSHKPDDEVNPELEPDQVYKFSVPVEHYVK YDNIIFPYERKEYILAFKNGYGVAAVNSKGEVYIGLITBESSGNEICGPIYCDKNK AAQAIRPISITIDLPMLKLYLVKCSDFEDSATWIRPPIRRSHENPFAHIFNRYI PDGLNFSKVFVSSNPNAPC"
CDS	
gene	complement(join(12240..12394,12467..12541,12691..12784,13262..13392,13465..13524,13876..14032)) /gene="F22B5.6" complement(join(12240..12394,12467..12541,12691..12784,13262..13392,13465..13524,13876..14032)) /codon_start=1

```

/protein_id="CAA90358.1"
/db_xref="GI:3876230"
/translation="MKNPYGFAYVNSCLSTWISOKRPSYFLALLLPVPEESMTL
LAOYHIHSSNITISMNLSRKIVENMANTHGLCTVATLCTLSLIYDIL
FTHTFRIVFNMIAVAVACGLFVGLVETLGLVETLGLVETLGLVETLGLVETL
NSTGEVRLESAGECTAISSEKVFPELLIOAVIFLPHILLTFMSWFSGLYPRM
HEDFE"
join(15384..15575,15622..16479,16527..17105,18608..18694,
18794..18946,18996..19449,19497..19843,20042..20554,
20672..21600,21677..21812)
/gene="F22B5.7"
join(15384..15575,15622..16479,16527..17105,18608..18694,
18794..18946,18996..19449,19497..19843,20042..20554,
20672..21600,21677..21812)
/feature="F22B5.7"
/note="limited similarity with the elongation factor EF-3
(Swiss Prot accession number P25997)"
CDNA EST EMBL:189075 comes from this gene
CDNA EST YK2848.3 comes from this gene
CDNA EST YK2848.3 comes from this gene
CDNA EST YK15387.5 comes from this gene
CDNA EST YK420b2.5 comes from this gene
CDNA EST YK421d3.3 comes from this gene
CDNA EST YK421d3.3 comes from this gene
CDNA EST YK396b3.3 comes from this gene
CDNA EST YK396b3.3 comes from this gene
CDNA EST YK396b3.3 comes from this gene
CDNA EST YK359g5.3 comes from this gene
CDNA EST YK359g5.3 comes from this gene
CDNA EST YK351a5.5 comes from this gene
CDNA EST YK351a5.5 comes from this gene
CDNA EST YK351a1.3 comes from this gene
CDNA EST YK351a1.5 comes from this gene

Query Match      29.2%: Score 339.2; DB 3; Length 25823;
Best Local Similarity 63.3%; Pred. No. 4.5e-69;
Matches 633; Conservative 0; Mismatches 233; Indels 134; Gaps 2;

QY 116 CAATCAATTTTAATACAGTATGATGCAAGACATGTTATGTGAAGTCTTATG 175
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4051 CAGTAACTTCAACACAGCTAATGCTTGCAGGAGACAGCTTACGTTAAGAGACTTTTCG 3992

QY 176 ATCAAGAAGGTTCGCGTATGATGAGGTGACGCTCAAGTTGCCGAATTTCACTTCAT 235
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3991 ATCAACAAGAAATGAGAAATACGAGAGAGGTGCTCAAGTCTGCAATCAACTTCAT 3932

QY 236 TTGATTCATGCAATGTTGCGGCTACAGATCTGCAATCCAGCGTATTTTGTACAA 295
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3931 TCGATACCTGCAAGTTCGCTGACAGATCTGCAATCCAGAGAGTCTTCCTTACCA 3872

QY 296 CAACGTGTCATGCTGTTTCATCAATTAATTTGTACCAAGTTGATC----- 343
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3871 CAACGTGTCGCTGCTTCATCCCAATTTGTACCAAGTTGATCGGTAGACATATG 3812

QY 344 ----- 343

DB 3811 AAACCTGTAATTTTAAGAAATGCTTCTGTAAGTTAGATACGAATTAAGAGC 3752

QY 344 -----GTGCATATCG 353

DB 3751 TTGATATCACTAGAGAGCTTCTGTTAAATAAATGCAATTTGAATTTCCAGTCCCTACCG 3692

QY 354 AGTACATAGCTTTTACATGAGAGCTGATTAACAAGTTAGTACAGATGAGTATCTGA 413
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3691 TGTTCATGCTTACATGAGAGCCGACAGACAGTCTTACCAATTAAGTCTCTGA 3632

QY 414 AATCAACAATGCTTTTCAACCAATTTGCGCGAGTACAGTATGCGGTTATGAATTTT 473
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3631 TCTTACACGCGTTTCCAAACCAAGTGTTCCTCAATGCTCAATTTACAGATTTCT 3572

QY 474 GGTGATGACCAACCGGTCACCAATTTGCTATCATTTGTCAGACAGTATATCA 533
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

DB 3571 TAACGAGAGACCACTGAGAAACAGTTCCAACTATTTGACAACAAGTCTATCA 3512
QY 534 TAAATGACATGCGATTTGAAACCGTGTATCTTTCGCGGGTGTTCATCTCTGCTT 593
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3511 CAAATGAGACCTGTATTCGAAACCGTGTATCTTTCGCGGGTGTTCATCTCTGCTT 3452
QY 594 TGTGATGATGTTAAGCGTATGATCTGTGAAATTTGATGCTGATGATGCTCTTGA 653
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3451 TGTGATGATGATGAAATGATGATACCGTCAATTTTGTGATGAGAACGTTTGCTCTTGA 3392
QY 654 TAAATATTTGCTAATTAATTTGGAATATTCACAGATTTAATGCTGCGCAAGAGCTCA 713
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3391 TAAATTTGCTCAACAAATTTGGAATATTCACAGATTTAATGCTGCGCAAGAGCTCA 3332
QY 714 GGTATCAAAATATGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 773
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3331 GGTATCAAAATATGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3272
QY 774 TAAAGAACCAATATGACCAATGTTTCGACCAAAATGTTTCAGAACCAAGAGATTGCGAGC 833
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3271 TAAAGAACCAATATGAGAGTGTGCTGCGCCCAACGCTGTGAGCCAAAGAGATTGAGC 3212
QY 834 TGTAAACAGGTGTCGCGCAGCAAAACCTGTCGAGCTGCGCAACTTGTCTTACTCAA 893
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3211 AGTTAA-----GCAAGCTAACCAAACTGCTGAGTTCTTCAGAGTTTGAAGAA 3164
QY 894 GAAAGATCTGACAGAACCGAATATCATTTATGATGATGATGATGATGATGATGATGATGAT 953
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3163 GAGATCCGCTCCAGTAATGAGAAATATTTCTGATGCTCGGCTGAGCTCAACACTTTTGA 3104
QY 954 AATTAGGATGATATATCAAGCTTTCAGTTGATTTACGT 993
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3103 AGTTTGAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3064

RESULT 11
CEC47G2
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REMARK
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Submitted (27-Apr-1995) Nematode Sequencing Project, Sanger
Institute, Hinxton, Cambridge CB10 1SA, England and Department of
Genetics, Washington University, St. Louis, MO 63110, USA. E-mail:
jes@sanger.ac.uk or rwenematode.wustl.edu
Coding sequences below are predicted from computer analysis, using
predictions from GeneFINDER (P. Green, U. Washington), and other
available information.
Current sequence finishing criteria for the C. elegans genome
sequencing consortium are that all bases are either sequenced
unambiguously on both strands, or on a single strand with both a
dye primer and dye terminator reaction, from distinct subclones.
Exceptions are indicated by an explicit note.
IMPORTANT: This sequence is not the entire insert of clone C47G2.
It may be shorter because we only sequence overlapping sections
once, or longer because we arrange for a small overlap between
neighbouring submissions.

```


FEATURES
source

The true left end of clone C47G2 is at 1 in this sequence. The true right end of clone C47G2 is at 6388 in sequence 249912.
The true left end of clone T24P1 is at 34731 in this sequence. The true right end of clone T05B9 is at 12750 in this sequence. The start of this sequence (1..101) overlaps with the end of sequence 249129.
The end of this sequence (34731..34831) overlaps with the start of sequence 249912.
For a graphical representation of this sequence and its analysis see: <http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=C47G2>
IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.
Location/Qualifiers
1..34831
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="II"
/clone="C47G2"
/join(2562..2679,2992..3217,3898..4636,4687..4878)
/gene="C47G2.1"
/join(2562..2679,2992..3217,3898..4636,4687..4878)
/gene="C47G2.1"
/note="similar to cuticulin"

gene
CDS

CDNA EST YK146a2.5 comes from this gene
CDNA EST YK146a2.3 comes from this gene
CDNA EST YK262c12.3 comes from this gene
CDNA EST YK262c12.5 comes from this gene
CDNA EST YK265b3.5 comes from this gene
CDNA EST YK27898.3 comes from this gene
CDNA EST YK27898.5 comes from this gene
CDNA EST YK439e8.3 comes from this gene
CDNA EST YK439e8.5 comes from this gene
CDNA EST YK380a11.3 comes from this gene
CDNA EST YK583c11.3 comes from this gene
CDNA EST YK380a11.5 comes from this gene
CDNA EST YK583c11.5 comes from this gene
/codon_start=1
/protein_id="CA88934.1"
/db_xref="GI:3875030"
/db_xref="SWISS-PROT:Q03755"
/translation="MTWKPIICIAALVAVNPNVNEGEPEVCEGPNSTIVNT
RNFPEHYVTKGIDYDAGCRSDGSGQVAGCIEPRSCMTAPRSLNPGVSTTV
ISFHPQVTKVDPAIRIQCFRMSDKTSTQIETVSQITTAFTQVYPMYCKEITIDG
KPSGQPIQFATIGQVYHAWTKDSEITDIFCAVSHCTVDNGDVIQLNEGCLAD
KFLNNLEPTDLMAGQEAHVYKADRSOLFQCOISITIKDGSFCARPTSEPOGF
GAVKQAGAGAHAAAPDAQAGVEVQAAPVAAAPVAAAPVAAAPVAAAPVAAAPVAA
LRKRSFGENEGILDAVEINTLIDMEGASPSAPAAALVSESVRRRATSGICILTP
IGFASFLIGTITATLSATIEFVARPTSHKH"

gene
CDS

join(15433..15770,16205..16531,16772..17108)
/gene="C47G2.2"
/note="contains similarity to Pfam domain: PF00250 (Fork head domain). Score=22.2, E-value=2.5e-63, N=1
CDNA EST EMBL:T01618 comes from this gene
CDNA EST YK123h2.5 comes from this gene
CDNA EST YK123h2.3 comes from this gene
CDNA EST YK268f8.3 comes from this gene
CDNA EST YK268f8.5 comes from this gene
CDNA EST YK474g11.5 comes from this gene
CDNA EST YK474g11.3 comes from this gene
CDNA EST YK670a12.3 comes from this gene
CDNA EST YK637f8.5 comes from this gene
CDNA EST YK670a12.5 comes from this gene
/codon_start=1
/protein_id="CA88935.1"
/db_xref="GI:3875031"
/db_xref="SPTREMBL:Q18694"
/translation="MLFSMESILSTIKLEPPPKLEPVITNEQVVDLPRNSTRLSE

gene
CDS

PTSSAVLEHDLKFGESRRRSRLGDEPTEDEGVPYRKANKRNHSTSSAADSSDDA
KDDDDDDSTSRKSMGSRKSHAKPPYSYALIAMLSLSEPKLITSECEFIHK
FEYKREKPAWQNSIRHNLSDCFVYAKPGPCPGNGNVALDPNCEMDNCSFIR
RRKRYKKNSDYHEHMSHPPEPPELPQCGPPEPRNMHPANLPIGLHPNPAVYN
MPAFIIPQNDISQILSMASRIRPMQAPVSSGQKRTSSSSSPNENSSAVSDKLSA
complement(join(22244..22507,23220..23597))
/gene="C47G2.3"
complement(join(22244..22507,23220..23597))
/gene="C47G2.3"
/codon_start=1
/protein_id="CAB97233.1"
/db_xref="GI:9367121"
/db_xref="SPTREMBL:Q9A09"
/translation="MALASSNATSWIEELFGNPPFRKKERYVEEKPPEYPSAYQMI
DOMGIKTRPNPPTPIKPIOMLIPMSREERIMQMGEMCTKASISGLGVG
FAGFLTASVDPQSLMVGDDPTKQITLKQTKRKESRRKSYGKNFGSLMGFSGTECA
LETTIRAKSDWRNGYSGGIVGLGLGRNGYAPVWGAAGPVESTIIDHYKRC"
complement(join(24417..24975,25043..25176,25237..25522,
25668..25907,25967..26054,26106..26227,26277..26608,
26681..26785,26846..26914))
/gene="C47G2.4"
complement(join(24417..24975,25043..25176,25237..25522,
25668..25907,25967..26054,26106..26227,26277..26608,
26681..26785,26846..26914))
/gene="C47G2.4"
/note="similarity to Trichostrongylus colubriformis 11 kd
secretory protein (Swiss Prot accession number P21937)
CDNA EST YK299t.3 comes from this gene
CDNA EST YK299t.5 comes from this gene
CDNA EST YK6d5.5 comes from this gene
CDNA EST YK179h8.5 comes from this gene
CDNA EST YK179h8.3 comes from this gene
CDNA EST YK358b10.3 comes from this gene
CDNA EST YK358b10.5 comes from this gene
CDNA EST YK301f4.3 comes from this gene
CDNA EST YK301f4.5 comes from this gene
CDNA EST YK295f2.3 comes from this gene
CDNA EST YK295f2.5 comes from this gene
CDNA EST YK550d1.3 comes from this gene
/codon_start=1
/protein_id="CA88936.1"
/db_xref="GI:3875032"
/db_xref="SPTREMBL:Q18695"
/translation="MGTISLAVQLTFVFLNLSYLKSTIRKQNPITYTISTFGHYF
SLIIVFVLADVAITFFPKCENDRQRIINTTSTPAVPECELPGGVPPDVLENLM
RVVYSAGQLILIPLOSQYTAGNEFLIGIRAVANINMLYVAYSLCEALITVA
MRKGYINTEINKVIVSASNTWGLFLLVLLGHLELPSLIMHGNRHYRLKTYF
DIEXLASEKSEAEENVKDYYKRVYLFPMKMKDSGQRKVTILSKRSDVIDNLF
SHOVIDNALHDSGCSSEAKLTLIKKRTIYANQIANNATQAKVLDALFLENLAFS
ESNGYNLELSRNTGVPFGRFRWYTRLOTPRCRIIGIYVPEKTPFVLESECTFVSY
TLSPAFTVEISVSTFRPKYTYTFVAFGLIYVLTISATYFTRLOTKYIYHDNPHTD
ENSILFSAIILCRTPICLNLHGMIMHDSHSMKSFGEIYQFKLGHLDVFTILA
KQINILYPIICILCAIHYRGAVALYHIMHDSHSMKSFGEIYQFKLGHLDVFTILA
SIKRSNDROSPRTQMTNRSFGSSNAGSGTTSFKRSNNKNDERRPMLDEDDVEBSST
RISKMSPTREPPSSSGPFDMM"

gene
CDS

join(30076..30134,30184..30492,31050..31274,31321..31480,
31534..31653,31701..31835,31889..32104,32160..32570,
32639..32935,32998..33215,33263..33651,33802..34049)
/gene="C47G2.5"
join(30076..30134,30184..30492,31050..31274,31321..31480,
31534..31653,31701..31835,31889..32104,32160..32570,
32639..32935,32998..33215,33263..33651,33802..34049)
/gene="C47G2.5"
/note="CDNA EST EMBL:M79633 comes from this gene
CDNA EST YK179g3.5 comes from this gene
CDNA EST YK227d8.5 comes from this gene
CDNA EST YK257e7.5 comes from this gene
CDNA EST YK319c4.5 comes from this gene
CDNA EST YK331d4.5 comes from this gene
CDNA EST YK335b12.5 comes from this gene
CDNA EST YK486d7.5 comes from this gene
CDNA EST YK513d7.5 comes from this gene
CDNA EST YK615h7.5 comes from this gene"


```

/codon_start=1
/protein_id="CAAB8937.1"
/db_xref="GI:3875033"
/db_xref="SPTREMBL:Q18696"
/translaton="MMAEGRKRRNNKRSRRDREGTSASDSAYSDSDAEDITQDDDD
PMLNLDSLDSDHNVAREPEQEPDLNVHSPNMEQENODLIEHGKAFPEQSGSR
IAOESASVATVDGRSDSDSHATRSKLTHYSKRSMPKLTSMKSTSTFOEWMKEE
ENSLRLKLTQNTFPLEDLNBPVVOESRKAELVYVTSREMKALLETSLMPKIN
TDLPMKQOYRLSTASELTLTTRGDIWPEKQOYVTEPERKCYLDLNDKPTLNHLYAGE
FARKIELCSHREFDLTSLDQETKFDCKLNINIGALELLENLVRIPTTSEGRIV
KEMWISLENFPEKIVDMRKESTEDDEKELAVELYELIEMDKYITMSKYDHLHARS
MDELTINIAQNLIVEEQSPAEPELVKCKSALISASKILAEIKTNFVSNAAQOLEE
ERNLSEHRYSYGLMRPCMDNDPEYHSQOPPERIVEGILNRPNILOTYRIEAAE
GSVMQPLRLIIEICNTCMSTHKEINAVAFSLPFINIKRAKMI.PRSVAVHCLLVK
VILLHSSPCDELSPAEVLITTEGGLQNTYDPAISPNPSSVACSLRBFNONGE
AINRKAAGIPIKOKILATLSADNTWTELEDILIHILNKLHRQMDHFNDDSVSIRK
DSHGKNDSEEWTDASTFAEMDATSAAQAFSGSSPEPMMQARSDEQGFDDTPEE
DEFRLCSERANSSSCAIGSFETSPKIKMGAEKSTSEKASPPSVASTYPOQTNGMO

```

QY	330	TACCAAGTTATGTCGATATCGAGTACATGCTTTTACATGGAAGCTAAAAACGT	389
Db	3883	TTCTAAATTTCCAGCGCCTCACCGTATCCAAATGCTTCTACATGAGTCCGACAAACTGT	3942
QY	390	TATGTCACAGATTGAGTATCTGAAATACCAACTGCTTTTCAAACTCAAAATTTGCCGAT	449
Db	3943	GTCCACACAGATTCAGAGCTCGGATCTCCACCCTCTCTCCAGACCAGGTGTACCAAT	4002
QY	450	GCCAGTATGCCGTATGAAATTTTGGATGTGGACCAACCGGTCAACCAAGTTCAATTTGC	509
Db	4003	GCCAGTCTGCAAAATACGAGATCTTGGACGAGGACCATCCGACACCAATTCAGTTGCG	4062
QY	510	TATCATTTGTCAGCCAGTTTATCATTAATATGGAACATCGATTTCTGAACCGTTGATCTTT	569
Db	4063	CACCATGGGACAAACAAGTCTATCAACAATATGACTTGGCATTTCTGAGACACTGACACTTT	4122
QY	570	CTGCGCGGTTGTCATATTCCTGCTTGTTCGATGATGATGATACGGTGTACTGTGGAATTC	629
Db	4123	CTGCGCGGTCCTTCACTCTTCCACTGTTCGATGATGATGATGATGATGATGATGATGAT	4182
QY	630	AAATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	689
Db	4183	TATGGAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	4242
QY	690	TTTAAATGCTGCGCCAGAGCTCACGATATACAAATATCGGATCGATACAGCTTTTCTA	749
Db	4243	CTTATGCTGCTGCGCCAGAGCTCACGATGATGATGATGATGATGATGATGATGATGATGAT	4302
QY	750	TCAATGCCAGATCATGATATTCATTAATAAACAACAATATGCGATGATGATGATGATGATGAT	809
Db	4303	TCAATGCCAGATCATGATATTCATTAATAAACAACAATATGCGATGATGATGATGATGATGAT	4362
QY	810	TTTGAACCAACAAGATTCGAGGCTGTTAAACAAGATGATGATGATGATGATGATGATGATGAT	869
Db	4363	CTCAGAGCAACAAGATTCGAGGCTGTTAAACAAGATGATGATGATGATGATGATGATGATGAT	4422
QY	870	AGCTGCGCAAC 880	
Db	4423	CGCTGCTCCAC 4433	

RESULT 12

CELCUT1

LOCUS

DEFINITION

VERSION

KEYWORDS

SOURCE

ORGANISM

CELCUT1

C.elegans cuticlin 1 (cut-1) gene, 3' end.

M55997.1 GI:156271

cuticle protein; cuticlin 1.

C.elegans DNA.

Caenorhabditis elegans

2310 bp

DNA

linear

INV 26-APR-1993

REFERENCE	THORS	FILE	JRNAL	LINE	JBMED	URES	source
Eukaryota; Metazoa; Nematoda; Chromodorea; Rhabditids; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.							
1 (bases 1 to 2310)							
Sebastiano, M., Lassandro, F. and Bazicalupo, P.							
cut-1 a Caenorhabditis elegans gene coding for a dauer-specific noncollagenous component of the cuticle							
Dev. Biol. 146 (2), 519-530 (1991)							
91323673							
1864469							
							Location/Qualifiers
							1. .2310
							/organism="Caenorhabditis elegans"
							/strain="N2"
							/db_xref="taxon:6239"
							join(1186. .1920,1971. .2310)
							/gene="cut-1"
							join(1186. .1920,1971. .2310)
							/partial
							/gene="cut-1"
							/product="cuticle1n 1"
							join(1186. .1920,1971. .2162)
							join(1186. .1920,1971. .2162)
CDS							

[illegible]

```

Db 1558 TACAATATGCGATGCTCCCACTCTCTATTCATGCGCAATATCCATCCATCAACCAAG 1617
Qy 778 GAACCAAAATAGGAATGTTTGCACACATGTTAGAACCAAGAGATTGCGAGCTGT 837
Db 1618 GACCCAGGAAGGAATGTCGCCGCTCCAACTGCTCAGACACCAAGAGATTGCGAGCGTC 1677
Qy 838 AAAACAGGTGTGCGCGCAAAACCGTGTGCGACGTCGCGCAAC 880
Db 1678 AAACAAGCTGTGCGCGAGAGACTATGCGCGCTGCTCCAC 1720

RESULT 13
LOCUS AR204524 813 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 8 from patent US 6368600.
ACCESSION AR204524
VERSION AR204524.1 GI:21501872
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 813)
AUTHORS Chandrasekar, R. and Morales, T.H.
TITLE Parasitic helminth cuticlin nucleic acid molecules and uses thereof
JOURNAL Patent: US 6368600-A 8 09-Apr-2002;
FEATURES
source Location/Qualifiers
1..813
BASE COUNT 247 a 148 c 164 g 254 t
ORIGIN

Query Match 26.7%; Score 310.4; DB 6; Length 813;
Best Local Similarity 70.7%; Pred. No. 2.8e-62;
Matches 455; Conservative 0; Mismatches 161; Indels 28; Gaps 2;

Qy 372 GGAAGCTGATAAAGACAGTATGTC-ACAGATTGAGTATGTAATGACAACTGCTTTTC 430
Db 14 GAAAGCTGATAAAGACAGTATGTC-ACAGATTGAGTATGTAATGACAACTGCTTTTC 430
Qy 431 AAATCAAAATTTGCCGATGCGCATATGCCGTTATGAATTTTGGATGGGACCAACCG 490
Db 74 AAATCAAGTGTATGACCAATCCCGTATGTCATGATGATTTGGAAGTGGACCAACTG 133
Qy 491 GTCACACAGTTTCAATTTGCTATCATGTCGTCAGCGATTTATCAATTAATGACATG 550
Db 134 GTGACCTGTTCATTTGCAATGATTTGAGATCATGTATATCACAATGACATGATTT 193
Qy 551 CTGAAACCGTTGATCTTCTGCGCGGTTGCCATTCCTGCTTGGATGATGATGTAACG 610
Db 194 CAGAGCTACAGATACATCTCTGTCATTTAGTACATTCATGTTGGATGATGAGAAAG 253
Qy 611 GTATACCTGTGGAATTTCTAAATGCTGATGATGCTCTTGAATTAATATTGCTAAATA 670
Db 254 GTATGCTAGTGAGATTTCTAAATGAGAGATGCTTTGGACAAATATTACTCAATA 313
Qy 671 ATTGGAATATCCACAGATTTAATGCTGGCCCAAGAGCTCACGATTAATAATATGCGG 730
Db 314 ATTGGAATATTTACAGATTTAATGCTGGCCCAAGAGCTCATGTTTAATAATATGCG 373
Qy 731 ATGATCAGACGTTTCTATCAATGCCAGATCAGTATTACATTAAGAACAATATGCG 790
Db 374 ATGATCAGACGTTTCTATCAATGCCAGATTTAGTATTAACAATTAAGACCCACATGCG 433
Qy 791 AATGTTGACACCAATGTTCAAGAACCAAGAGATTGCG- 830
Db 434 AATGTTGCTGACACCAATGACAGACCAAGAGATTGCGCATTAATATCGACAG 493
Qy 831 -----ACGTTTAAAGAGTGTGCGCGACCAAAACCTGCTGACGTCGCAACTTC 883
Db 494 GATTGCTCTGTAAATATGCTGCTGACACACAGCTCCAGAGAGCTCTTCTTCCAC 553
Qy 884 GTTACTCAAGAAAGATGTCGAGAACCGGAGATATCATTTGATGACAGACTGATATCA 943

```

```

Db 554 GATTATACAGAGGATCAATTAATTTCTGATAAATACGCTGAGCTCAGTACCGGTTTAA 613
Qy 944 ACACCTTGAATTTAGCGATGATTAATCAAGCTTTCAGGTTGAT 987
Db 614 GCACGGTGTATTAATCAACCAAGAGATCCGAACCTTCTCAGCAAT 657

RESULT 14
LOCUS AR204525 813 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 10 from patent US 6368600.
ACCESSION AR204525
VERSION AR204525.1 GI:21501874
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 813)
AUTHORS Chandrasekar, R. and Morales, T.H.
TITLE Parasitic helminth cuticlin nucleic acid molecules and uses thereof
JOURNAL Patent: US 6368600-A 10 09-Apr-2002;
FEATURES
source Location/Qualifiers
1..813
BASE COUNT 254 a 164 c 148 g 247 t
ORIGIN

Query Match 26.7%; Score 310.4; DB 6; Length 813;
Best Local Similarity 70.7%; Pred. No. 2.8e-62;
Matches 455; Conservative 0; Mismatches 161; Indels 28; Gaps 2;

Qy 372 GGAAGCTGATAAAGACAGTATGTC-ACAGATTGAGTATGTAATGACAACTGCTTTTC 430
Db 800 GAAAGCTGATAAAGACAGTATGTC-ACAGATTGAGTATGTAATGACAACTGCTTTTC 430
Qy 431 AAATCAAAATTTGCCGATGCGCATATGCCGTTATGAATTTTGGATGGGACCAACCG 490
Db 740 AAATCAAGTGTATGACCAATCCCGTATGTCATGATGATTTGGAAGTGGACCAACTG 681
Qy 491 GTCACACAGTTTCAATTTGCTATCATGTCGTCAGCGATTTATCAATTAATGACATG 550
Db 680 GTGACCTGTTCATTTGCAATGATTTGAGATCATGTATATCACAATGACATGATTT 621
Qy 551 CTGAAACCGTTGATCTTCTGCGCGGTTGCCATTCCTGCTTGGATGATGATGTAACG 610
Db 620 CAGAGCTACAGATACATCTCTGTCATTTAGTACATTCATGTTGGATGATGAGAAAG 561
Qy 611 GTATACCTGTGGAATTTCTAAATGCTGATGATGCTCTTGAATTAATATTGCTAAATA 670
Db 560 GTATGCTAGTGAGATTTCTAAATGAGAGATGCTTTGGACAAATATTACTCAATA 501
Qy 671 ATTGGAATATCCACAGATTTAATGCTGGCCCAAGAGCTCACGATTAATAATATGCGG 730
Db 500 ATTGGAATATTTACAGATTTAATGCTGGCCCAAGAGCTCATGTTTAATAATATGCG 441
Qy 731 ATGATCAGACGTTTCTATCAATGCCAGATCAGTATTACATTAAGAACAATATGCG 790
Db 440 ATGATCAGACGTTTCTATCAATGCCAGATTTAGTATTAACAATTAAGACCCACATGCG 381
Qy 791 AATGTTGACACCAATGTTCAAGAACCAAGAGATTGCG- 830
Db 380 AATGTTGCTGACACCAATGACAGACCAAGAGATTGCGCATTAATATCGACAG 321
Qy 831 -----ACGTTTAAAGAGTGTGCGCGACCAAAACCTGCTGACGTCGCAACTTC 883
Db 320 GATTGCTCTGTAAATATGCTGCTGACACAGCTCCAGAGAGCTCTTCTTCCAC 261
Qy 884 GTTACTCAAGAAAGATGTCGAGAACCGGAGATATCATTTGATGACAGACTGATATCA 943
Db 260 GATTGATCAAGAGAGATTAATTTCTGATTAATACGTTGAGAGCTGACGATACCGGTTTAA 201
Qy 944 ACACCTTGAATTTAGCGATGATTAATCAAGCTTTCAGAGTTGAT 987

```

Db 200 GCACGGTTGATATACCGAGAGAAATCCGAACTCTCAGAAAT 157

111 1111 11 1111 11 1111 11 1111 11

RESULT 15

ALU73005

LOCUS

DEFINITION

ALU73005 5792 bp DNA linear INV 14-AUG-1997

Ascaris lumbricoidea CUT-1-like cuticlin protein precursor

(ascut-1) gene, complete cds.

ACCESSION

U73005

VERSION

U73005.1 GI:1657624

KEYWORDS

SOURCE

Ascaris lumbricoidea.

ORGANISM

Ascaris lumbricoidea.

Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida;

Ascaridoidea; Ascarididae; Ascaris.

REFERENCE

1 (bases 1 to 5792)

AUTHORS

Timinouni,M. and Bazzicalupo,P.

TITLE

cut-1-like genes of Ascaris lumbricoidea

JOURNAL

Gene 193 (1), 81-87 (1997)

MEDLINE

97390131

PUBMED

9249070

REFERENCE

2 (bases 1 to 5792)

AUTHORS

Timinouni,M. and Bazzicalupo,P.

TITLE

Direct Submission

JOURNAL

Submitted (01-OCT-1996) GA3, I.I.G.B., via G. Marconi, 10, Napoli, NA 80125, Italy

FEATURES

source

1. 5792

Location/Qualifiers

/organism="Ascaris lumbricoidea"

/variety="suum"

/db_xref="taxon:6252"

766..5081

/gene="ascut-1"

join(766..954,2851..3239,3583..4083,4499..5081)

/gene="ascut-1"

/product="CUT-1-like cuticlin protein ASCUT-1"

766..881

/gene="ascut-1"

766

/gene="ascut-1"

/note="site of transcribing by the SL1"

join(882..954,2851..3239,3583..4083,4499..4693)

/gene="ascut-1"

/note="ascut-1"

/note="cuticlin gene; ASCUT-1: homologous to cuticlin proteins of other nematodes"

/codon_start=1

/product="CUT-1-like cuticlin protein precursor"

/protein_id="FAB66646.1"

/db_xref="GI:1657625"

/translation="MCRVSVFLALFGLAAIPVDNGVEGEPEIECGPTSTIVFNTRN
PREGHAIKGLIDEGCRSDGKQVAGISLPDSCNVATKSLNPRGIVTTTVIS
FHLFTKVDRAIVQCCFYMEADRTVSTQIEVSEITTAFTQIIPMPVCREILDDGP
TQPIQPIATIGQOVVHKWCDSEVTDFCAVNSVCSGSGTPIIILNEGALDKY
LNNLEIYPTDLMAGQEAHYKYADRSQLEFYOCQITTIKEPNSECRPTCSFQGFCA
VREGSIAPIKORCOLRIKSGSDYDNTVDTPESALDISDRDEALPMDLRHAR
HARGQVITISPAINEGICMSPFGSIFMGLAVLAAVVVVSKLRPQKA"

882..929

/gene="ascut-1"

join(930..954,2851..3239,3583..4083,4499..4690)

/gene="ascut-1"

/product="CUT-1-like cuticlin protein"

4694..5081

/gene="ascut-1"

5064..5070

/gene="ascut-1"

BASE COUNT 1689 a 1254 c 1213 g 1636 t

ORIGIN

Query Match 21.4%; Score 249; DB 3; Length 5792;

Best Local Similarity 69.7%; Pred. No. 6,9e-48;

Matches 352; Conservative 0; Mismatches 150; Indels 3; Gaps 1;

QY 462 TTATGAATTTTGGATGGTGAGCCACCGGTCACCAAGTTCATTTGCTATTCATGGTCA 521

11 1111 11 1111 11 1111 11 1111 11 1111 11 1111 11

Db 3576 TTTACAGATTTCGACGAGAGTCCACGGAGACACCGCATATCCGACCATCGGTCA 3635

QY 522 GCCAGTTTATCATTAATGATGATCGATTTCTGAAAACCGTTGATCTTCTGCGGGTGT 581

11 1111 11 1111 11 1111 11 1111 11 1111 11 1111 11

Db 3636 GCAAGTATACCAAGAGGAGATTGCGAATTCGGAAGAGTGATACGTTCTGCTGTCGT 3695

QY 582 CCATTCCTGCTTTGTCGATGATGATGATGATGATGATGATGATGATGATGATGATG 641

11 1111 11 1111 11 1111 11 1111 11 1111 11 1111 11

Db 3696 TCACATCATGCTTCGTTGATGATGATGATGATGATGATGATGATGATGATGATG 3755

QY 642 ATGTCCTGCTGATTAATATTTGCTAATTAATTTGGAATATCCACATTAATAGGCTG 701

1111 1111 1111 1111 11 11 11 11 11 11 1111 1111 11

Db 3756 TTGTCCTGTCGAAATATCCGTCGAAACACCTGAGTACCCACGATTTGATGGCGG 3815

QY 702 CCAAGAGCTCAGCTATATCAATATGCGGATGATGATGATGATGATGATGATGATGAT 761

11111111 1111 1111111111 1111111111 11111111 11111111

Db 3816 CCAAGAGCGGACGCTGTCATCAATATGCGGACCATGATGATGATGATGATGATGATG 3875

QY 762 CAGTATACCATTTAAGAACCAATAGGATGATGATGATGATGATGATGATGATGATG 821

11 1111 1111 1111 1111 111111 1111 11 11 11

Db 3876 TACTATCATCATTAAGAGCCAAACAGCATGATGCCACCTTGCACTGATGAGCGCA 3935

QY 822 AGGATTCGAGCGCTTAAACAGCGTGGGCC---GCAAGCAAACTGCTGACGTCGCGCA 878

111111 11 111111 11111111 11 1111 11 1111

Db 3936 AGGATTTGGGGCTGTGCGTCCAGGTGGCTCCATAGCGCGCAAAAGCAGCGCGCTGCA 3995

QY 879 ACTTCGTTTACTCAAGAAAGATCTGCAGAACCGAGATATTCATGATGACCAACTGA 938

1111 11 1111 1111 11 11 11 11 11 11 11 11 1111

Db 3996 ACTCGGCTGATCAAAAAGATGGGGTGACTATGACACACCTTGATGACGACCGCA 4055

QY 939 TATCAACACCTTGAATTAGCGAT 963

111 1111 11 1111 111

Db 4056 CTTGACGCGCCTCGCATTTAGTGAAT 4080

RESULT 16

MAMTCUT1

LOCUS

MAMTCUT1 2584 bp DNA linear INV 25-NOV-1997

DEFINITION

M.arteiella Mtcut-1 gene.

ACCESSION

X96677

VERSION

X96677.1 GI:2648040

KEYWORDS

cuticle protein; cuticlin 1; Mtcut-1 gene.

SOURCE

Meloidogyne arteiella

ORGANISM

Meloidogyne arteiella

Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;

Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne.

REFERENCE

1 (bases 1 to 2584)

AUTHORS

De Giorgi,C., De Luca,F., Di Vito,M. and Lambert,F.

TITLE

Modulation of expression at the level of splicing of cut-1 RNA in the infective second-stage juvenile of the plant parasitic nematode Meloidogyne arteiella

JOURNAL

Mol. Gen. Genet. 253 (5), 589-598 (1997)

MEDLINE

97218031

PUBMED

9065692

REFERENCE

2 (bases 1 to 2584)

AUTHORS

De Giorgi,C.

TITLE

Direct Submission

JOURNAL

Submitted (15-MAR-1996) C. De Giorgi, Dipartimento di Biochimica e Biologia Molecolare, Via Orabona 4, 70126, Bari, ITALY

FEATURES

source

1. 2584

/organism="Meloidogyne arteiella"

/db_xref="taxon:42426"

/clone="puc(3000)"

/clone="mp19(1500)"

/clone="lmbda7-2(1)"

join(467..943,1139..1735,1825..2052)

/gene="Mtcut-1"

join(<467..943,1139..1735,1825..>2052)

/gene="Mtcut-1"

join(467..943,1139..1735,1825..2052)

CDS

[illegible]

Query Match	18.7%;	Score 217.2;	DB 3;	Length 2584;
Best Local Similarity	56.8%;	Pred. No. 2e-40;		
Matches 581;	Conservative	0;	Mismatches 243;	Indels 198;
				Gaps 2;

QY	51	TTGATATTCGGGTGACAAATGCTGTCCGAAGGTGACCAAAATTTGAAATGTGACCAATCTC	110
Db	529	TGCATTTCCAGTGTGACACACGGTGTGGGAAGCCGACGAGATTGAATGTGGGCCCAACTTC	588
QY	111	AATAAACAATCAATTTTAAATACAGCTAATGCATTGCAAGACATGTTTATGTCAAAGGCTC	170
Db	589	AATACCGGTAAATTTCAACACAGCAACCTTTTTCGAGGGGACGTTCTATGTAAAGGCTC	648
QY	171	TTATGATCAAGAAGGTGGCGGTGAATGATGAAGTGGACGTCAAGTTGCCGAATTTCACT	230
Db	649	CTTTGACACAGCCGGATGCCGTTGCGATGACACGGCCACAAAGTGGCGGGAATTGAGCT	708
QY	231	TCCATTTGATTCATGCACAAATGTTGCGGTAACAGATCTGAAATCCACAGTGATTTTGT	290
Db	709	TCCGTTGCAGCTTGCATATGTTGCCGCGCATGTGCT--GAACCOAAAGAGTGTGCT	765
QY	291	AACAAACAACGTGTGCTATTTCTGTTTATTCATTATTTGTGTACCAAAAGTTGATCGTGCTA	350
Db	766	CTCCACACAGGTGTACTCTCTTCATCCACAGTTGTGACCCAAAGTTGACCGGGCCTA	825
QY	351	TCGAGTACAAATGCTTTTACATGGAACGTATAAACAGTTAGTGCACAGATTGAGGATTC	410
Db	826	CCGATGCCAATGCTTGTATATGGAAGACAAAGAACCGTCTGGCAGCAGCTCGAGGTGTC	885
QY	411	TGAAATACACAATCTGCTTTTCAAACTAAATTTGTGCCGATGCCAGTATGCCGTTATG---	466
Db	886	CGAGATCACAAACAGATTCACAGACCAGAGGTGGTCCATAGCTGTGTCCAATATGAGGT	945
QY	467	-----	466
Db	946	AAGTAAGAAAGAAAGACGCTCACTCTTCAGTAAATATCATATGCAACTTTGATTTGGAT	1005
QY	467	-----	466
Db	1006	AAATCGAGTGGTTTATTTCAACATTTTGCCTTCATTACCAACAGCTTTAAAGTGTCAAA	1065

[illegible]

RESULT	17			
LOCUS	BMA012617			
DEFINITION	Brugia malayi cut-1 gene, partial.	637 bp	DNA	linear
ACCESSION	AJ012617			INV 29-JUN-1999
VERSION	AJ012617.1	GI:3858954		
KEYWORDS	cut-1 gene; cuticlin.			
SOURCE	Brugia malayi.			
ORGANISM	Brugia malayi.			
REFERENCE	Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;			
AUTHORS	Onchocercidae; Brugia.			
TITLE	1 (bases 1 to 637)			
JOURNAL	Lewis,E., Hunter,S.J., Tetley,L., Nunes,C.P., Bazzicalupo,P. and			
MEDLINE	Devaney,F.			
PUBMED	cut-1-like genes are present in the filarial nematodes, Brugia			
REFERENCE	pahangi and Brugia malayi, and, as in other nematodes, code for			
AUTHORS	components of the cuticle			
TITLE	Mol. Biochem. Parasitol. 101 (1-2), 173-183 (1999)			
JOURNAL	9933397			
FEATURES	10413052			
source	2 (bases 1 to 637)			
gene	Devaney,E.			
exon	Direct Submission			
mRNA	Submitted (06-NOV-1998) Devaney E., Veterinary Parasitology,			
	University of Glasgow, Bearsden Road, Glasgow, G61 1QH, UK			
	location/Qualifiers			
	1..637			
	/organism="Brugia malayi"			
	/db_xref="taxon:6279"			
	/clone="cut-1"			
	1..484			
	/gene="cut-1"			
	1..164			
	/gene="cut-1"			
	/number=1			
	join(<2..164,266..>484)			

```

CDS
    /gene="cut-1"
    join(<2..164,266..>484)
    /gene="cut-1"
    /codon_start=1
    /product="cuticlin"
    /protein_id="CA10074.1"
    /db_xref="GI:3858955"
    /db_xref="SPTREMBL:O96775"
    /translation="VISFHPILSTKYDRIRVOCFMEADKIVSTOLEVSEITTAQT
    QIVPMFVCRLEIDGPTQPIQFATIGPVPVHKWCDSEVDTFCVAVHSCFVDDGN
    GDVEILSADGALDKYLLNLEYP"
    165..265
    /gene="cut-1"
    /number=1
    266..484
    /gene="cut-1"
    /number=2

Intron
    174 a 136 c 134 g 193 t

BASE COUNT
ORIGIN
    174 a 136 c 134 g 193 t

Query Match
    16.4%; Score 190.6; DB 3; Length 637;
    Best Local Similarity 68.3%; Pred. No. 3.7e-34;
    Matches 334; Conservative 0; Mismatches 54; Indels 101; Gaps 1;

QY 303 TGTCAATTCGTTCAATTCATTAATTTGTTACCAAGTGATCGTATCGAGTACATG 362
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    1 TGTAAATTCCTTCATTCACATGTCGTCACAAAAGTTGATCGACATACGAGTGCAGTG 60

QY 363 CTTTTCACATGGAAGCTGATGAACAGTAGTGCACAGATTGAGGTATCTGAATCACAC 422
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    61 CTCTATATGGAAGCTGACAAAAGCGTCACACCCAGATGAGGTGCTGMAATCACAC 120

QY 423 TGCCTTTCACACCAATTTGCCGATGCCAGTATGCCGTTATGAA----- 468
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    121 TGCCTTTCACACCAATTTGCCGATGCCAGTATGCCAGTATGCTTTTGA 180

QY 469 ----- 468

Db 181 ATTTTGAATTCGATTTTATATACAAATTCAAAACATGAAAATCTATTATCA 240

QY 469 -----ATTTTGATGTTGGACCAACCGGTCAACAGTT 501
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    241 CATAGTAGATGTGCAAAATATTTTATGATTTTGGATGGTGGACCAACCGGACCAAT 300

QY 502 CAATTCCTTCATTCGTCACAGTTCATATCAATGACATGCCATTCGAAACCGTT 561
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    301 CAATTCCTTCATTCGTCACAGTTCATATCAATGACATGCCATTCGAAACCGTT 360

QY 562 GATACCTTTCGCGGTTGTCATTCCTGCTTTCGATGATGATGATGATGATG 621
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    361 GATACCTTTCGCGGTTGTCATTCCTGCTTTCGATGATGATGATGATGATGATG 420

QY 622 GAATTCCTTAATGCTGATGATGATGCTTTCGATTAATATTTGCTAAATATTTGGAATAT 681
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    421 GAATTCCTTAATGCTGATGATGATGCTTTCGATTAATATTTGCTAAATATTTGGAATAT 480

QY 682 CCAACAGAT 690
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    Db 481 CCGACAGGT 489

RESULT 18
BPA012618 358 bp DNA linear INV 29-JUN-1999
LOCUS BPA012618
DEFINITION Brugia pahangi cut-1 gene, partial.
ACCESSION AJ012618.1 GI:3858956
VERSION AJ012618.1
KEYWORDS cut-1 gene; cuticlin.
SOURCE Brugia pahangi.
ORGANISM Brugia pahangi
Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
Onchocercidae; Brugia.
REFERENCE
1 (bases 1 to 358)

```

```

AUTHORS
    Lewis E., Hunter, S.J., Tetley, L., Nunes, C.P., Bazzicalupo, P. and
    Devaney, E.
TITLE
    cut-1-like genes are present in the filarial nematodes, Brugia
    pahangi and Brugia malayi, and, as in other nematodes, code for
    components of the cuticle
JOURNAL
    Mol. Biochem. Parasitol. 101 (1-2), 173-183 (1999)
MEDLINE
    99339397
PUBMED
    10413052
REFERENCE
    2 (bases 1 to 358)
AUTHORS
    Devaney, E.
TITLE
    Direct Submission
JOURNAL
    Submitted (06-NOV-1998) Devaney E., Veterinary Parasitology,
    University of Glasgow, Bearsden Road, Glasgow, G61 1QH, UK

FEATURES
    source
    1..358
    /organism="Brugia pahangi"
    /db_xref="taxon:6280"
    1..356
    /gene="cut-1"
    join(<1..191,350..>356)
    /gene="cut-1"
    join(<1..191,350..>356)
    /gene="cut-1"
    /note="unnamed protein product"
    /codon_start=1
    /protein_id="CA10075.1"
    /db_xref="GI:3970656"
    /db_xref="SPTREMBL:O96776"
    /translation="PIQFATIGPVPVHKWCDSEVDTFCVAVHSCFVDDGNDEI
    LNDGCGALDKYLLNLEYPFDL"
    1..191
    /gene="cut-1"
    /number=1
    192..349
    /gene="cut-1"
    /number=1
    350..356
    /gene="cut-1"
    /number=2

Intron
    192..349
    /gene="cut-1"
    /number=1
    350..356
    /gene="cut-1"
    /number=2

exon
    110 a 64 c 64 g 120 t

BASE COUNT
ORIGIN
    110 a 64 c 64 g 120 t

Query Match
    13.5%; Score 156.6; DB 3; Length 358;
    Best Local Similarity 87.7%; Pred. No. 3.6e-26;
    Matches 171; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 496 CCAATTCATTTGCTATCATGTCGACCAAGTTATCATTAATGACATGCGATTGAA 555
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    1 CCAATTCATTTGCTATCATGTCGACCAAGTTATCATTAATGACATGCGATTGCGAA 60

QY 556 ACCGTTATCTTTCGCGGTTGTCATTCCTGCTTTCGATGATGATGATGATGAT 615
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    61 ACCGTTATCTTTCGCGGTTGTCATTCCTGCTTTCGATGATGATGATGATGATG 120

QY 616 ACTGTGAAATTCCTTAATGCTGATGATGATGCTTTCGATTAATATTTGCTAAATATTTG 675
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    121 ACGGTGAAATTCCTTAATGCTGATGATGATGCTTTCGATTAATATTTGCTAAATATTTG 180

QY 676 GAATATCCACAGAT 690
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    Db 181 GAATATCCACAGAT 195

RESULT 19
CEP53FL 39478 bp DNA linear INV 11-DEC-2001
LOCUS CEP53FL/C
DEFINITION Caenorhabditis elegans cosmid F53FL, complete sequence.
ACCESSION Z81088.1 GI:1627965
VERSION Z81088.1
KEYWORDS HTG.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;

```



```

EHFESPKELMEVWRMPMTGYNKFATKEITKTDKMAAGLPIRKKDASHPLNNPAGAV
ROGAYIPHLRKLPGGSGAPSPAAPPGNONORPAPRANGNANPQPPPOOS
BOERKAFOLKKKEELIKVORANGDOLOPOMMEIORENEITSELSKTTI"
join(6082..6250,6307..6379,6432..6509,6564..6696,
6757..6930,7003..7053)
/gene="E04D5.1b"
join(6082..6250,6307..6379,6432..6509,6564..6696,
6757..6930,7003..7053)
/gene="E04D5.1b"
/note="cDNA EST yk656h4.5 comes from this gene"
/codon_start=1
/protein_id="CAC42280.1"
/db_xref="GI:14530389"
/translation="MPAPNTTSPRDSNEFVYSSFSNFLLIPKPKTFSPERLSOTLI
OIKMGDNLVYARSSSEGYLKRGLKDAVTEQKTSADVACNPAFVNNQOLPAP
CNOQVTRVEIATNKEITLCEVLEKTRKRLTLESPDNLTTEPNAVYGPKTKSQNQC
ESTVLRNESTYQSLOLQKHRSLSFRRMSRWLOEWLVAKVSFIQICLSTVITTSLSR
RVPTSH"
complement(join(8619..8761,8813..8988,9895..10028,
10089..10283,10330..10614,10730..10935,11582..11629,
11691..11718))
/gene="E04D5.2"
complement(join(8619..8761,8813..8988,9895..10028,
10089..10283,10330..10614,10730..10935,11582..11629,
11691..11718))
/gene="E04D5.2"
/codon_start=1
/protein_id="CAA91282.1"
/db_xref="GI:3875454"
/db_xref="SPTREMBL:019055"
/translation="MFGENLFVADHSTTPEKCEKIKIRVLESENQKHLFGFIKYP
INROLMAVRSLSPSSITPLATCDMLCTVTLTLEFGDADIVAKNTLTFVLR
YIYVFFIAKVOPAFPIILITTEFYIMTCERKRFGLTFVLTLSFRKFSKI
ENRGRIVIVFVCFEFSVALIRPVLVAKTEPLDDYRSSLGCTPEPAATEAEI
YDHVITAIOMPFVYVLLNLITIKRLAKRENNYPLIRGAGTTEVKKASFQO
NLPEENVYLOVADVIKESLIHRSSRSKROLNATYMTAIVTSYVCGVHFLTI
LEREDPSYLESTDRMQSSTFYVLSDPVSIQWMASSAIIPIYAKNPKLRQETIDY
IKREKSIETMSS"
join(25834..25873,27277..27512,27582..27721,27768..28110,
28158..28410,28491..28811,29092..29213)
/gene="E04D5.3"
join(25834..25873,27277..27512,27582..27721,27768..28110,
28158..28410,28491..28811,29092..29213)
/gene="E04D5.3"
/note="similar to cuticlin like protein"
/codon_start=1
/protein_id="CAA91280.1"
/db_xref="GI:3875452"
/db_xref="SPTREMBL:019053"
/translation="MPSRLAMKELKGGGPEVYCEASTLSLKRTRSNFGKVFQCY
VSEPSCTVGDGKTGRHFEVRHDSGVRORREINGVIAATVLIISRSITITIDRY
RVSCFVEGKRVNHNVDISALTQLESTQLPVCRYEILNAGSGPIKVAIGDQY
YAKMTCVAALENVYCMKHSCTVYDGGGPPVIDANGSGVGIIONLEYSDLA
GKLAPVFEKADKAGLYENCOILTIKDVNGCSNTOPOCTQSVYVPAQTKETAEI
YPYDSHESGTPTRPANYPVASSRPPIPTQAPASYPSPAPRPGADIDNGYDEPOPI
YIAEPENAYDGIIVGPNDEOPPTSAATQEDPVSRLLKRNVESIEQTCSTSRMFT
VYVGDIDPERGLIVFGLMEDEGETTNAGDHCAIIRLAREARNSQETQSTSRMFT
LILCLLFATVTVVFIYVOKORILAOIAFFKP"
join(29675..29840,31127..31214,31267..31348)
/gene="E04D5.4"
join(29675..29840,31127..31214,31267..31348)
/gene="E04D5.4"
/codon_start=1
/protein_id="CAA91281.1"
/db_xref="GI:3875453"
/db_xref="SPTREMBL:019054"
/translation="MMSINLTWHIILLVPSVSHVYTTAVTKATGETTVAGADILD
VSSSFYETTTASTCADPWTDCQTQYFLCSNKKYPIPLQOPKATKCGPGSGSTAP
VACVDVSTQ"
BASE COUNT 10458 a 5592 c 5636 g 9850 t
ORIGIN
Query Match 7.2%; Score 83.8; DB 3; Length 31536;

```

Best Local Similarity 56.1%; Pred. No. 4e-09;
Matches 179; Conservative 0; Mismatches 137; Indels 3; Gaps 1;

QY	480	TGACACACCGGTACACAGTTCAATTGCTATGTTGGTCAGCAGTTATCAATAATG	539
Db	27783	TGAAGCCGTCGATCACCATTAATAGTACGTCGATGAGATCAAGTATACCAATG	27842
QY	540	GACATGGCATCTGAACCCGTTGATCTTGTGGCGGTTGTCATTCCTCTTGGCGA	599
Db	27843	GACTTGGCTTCTGAACCTTGAATGCTATGATGATGAACTTACATATTTACGCTTGA	27902
QY	600	TGATGTAACTGATCTACTGTGAATTCGAA--ATGCTGATGATGATGCTTGTGATGA	656
Db	27903	TGATGTAACTGATCTACTGTGAATTCGAA--ATGCTGATGATGATGCTTGTGATGA	27962
QY	657	ATATTGCTTAATTAATTTGGATATTCACACAGATTAAATGCTGCGCAAGACTCACGT	716
Db	27963	GGTATCTCTTCAGATCTTGAATTCGATGATGATGATGATGATGATGATGATGATG	28022
QY	717	ATCAAAATATGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	776
Db	28023	CTTCAAAATTTGCTGACAAAGCTGCGCTTATTCATTAATTCATTAATTCATTAATCA	28082
QY	777	AGAACCAATATGCAATGT 795	
Db	28083	AGACGTAAATATGATGAT 28101	

RESULT 22
CEP53B6/C
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

CEP53B6 32412 bp DNA linear INV 26-APR-2002
Caenorhabditis elegans cosmid F53B6, complete sequence.
281086
281086.1 GI:1627952
HNG; Cuticulin; Human platelet tetraspan antigen like; Initiation factor associated protein; Thrombospondin like.
Caenorhabditis elegans.
Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
1
none.
Gene sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium
Science 282 (5396), 2012-2018 (1998)
99069613
PUBMED
9851916
REMARK
The C. elegans Sequencing Consortium.
2 (bases 1 to 32412)
AUTHORS
TITLE
JOURNAL
COMMENT

White, S.
Direct Submission
Submitted (21-OCT-1996) Nematode Sequencing Project, Sanger Institute, Hinxton, Cambridge CB10 1SA, England and Department of Genetics, Washington University, St. Louis, MO 63110, USA. E-mail: jess@angier.ac.uk or rwh@nematoe.wustl.edu
Coding sequences below are predicted from computer analysis, using predictions from GeneFinder (P. Green, U. Washington), and other available information.
Current sequence finishing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note.
IMPORTANT: This sequence is not the entire insert of clone F53B6. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.
The true left end of clone F53B6 is at 1 in this sequence. The true right end of clone F53B6 is at 5116 in this sequence.
The true left end of clone F32H2 is at 32306 in this sequence. The true right end of clone F08G11 is at 7984 in this sequence. The start of this sequence (1..105) overlaps with the end of sequence

280220.
The end of this sequence (32306..32412) overlaps with the start of
sequence 281523.
For a graphical representation of this sequence and its analysis
see: [http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?](http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=F53B6)
[name=F53B6](http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=F53B6)
IMPORTANT: This sequence is NOT necessarily the entire insert of
the specified clone. It may be shorter because we only sequence
overlapping sections once, or longer because we arrange for a small
overlap between neighbouring submissions.
Location/Qualifiers

FEATURES

source

1..32412
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="I"
/clone="F53B6"
join(1863..1946,2437..2628,2969..3154,3205..3308,
3430..3589,3804..3854)
/gene="F53B6.1"
join(1863..1946,2437..2628,2969..3154,3205..3308,
3430..3589,3804..3854)
/gene="F53B6.1"
/note="predicted using GeneFinder
Similarity to Human platelet-endothelial tetraspan antigen
(SW1PRT3_HUMAN), contains similarity to Pfam domain:
PF00335 (Transmembrane 4 family), Score=30.0,
E-value=5.8e-09, N=1"
/codon_start=1
/protein_id="CAB03120.2"
/db_xref="GI:11065632"
/db_xref="SPTREMBL:O9XVM9"
/translation="MGALDSAYGANGRIKSYVTALISILFISICICIMILAR
KQYALVSPSLYVDGRLVLIISLILNYLICEFALFKEMKCFVTCASVASIYAV
MLIGGICLGNRDLTHYTPNLKMLTSLRELYGHDKGJTESDALQSNFKCGV
NCTDNAQIKWTSKMYORAPKLLIPESCCISELERCSNPFDDAPPPYTSICYE
PLONDLHVMANSMCLTNATVOITIPSYAGCMYSKILRK"
complement(join(4270..4339,4493..4547,5071..5295,
5368..5662,5891..6148,6204..6752,6818..7287,7546..7676,
8104..8184,12576..12775,13246..13384,13931..14132,
14183..14294,14341..14457,14994..15155,15198..15311))
/gene="F53B6.2"
complement(join(4270..4339,4493..4547,5071..5295,
5368..5662,5891..6148,6204..6752,6818..7287,7546..7676,
8104..8184,12576..12775,13246..13384,13931..14132,
14183..14294,14341..14457,14994..15155,15198..15311))
/gene="F53B6.2"
/note="predicted using GeneFinder
contains similarity to Pfam domain: PF00047
(Immunoglobulin domain), Score=24.4, E-value=1.7e-06, N=1
CDNA EST EMBL:214404 comes from this gene
CDNA EST YK153a1.5 comes from this gene
CDNA EST YK195c9.3 comes from this gene
CDNA EST YK195c9.5 comes from this gene
CDNA EST YK286g7.3 comes from this gene
CDNA EST YK286g7.5 comes from this gene
CDNA EST YK323e3.5 comes from this gene
CDNA EST YK353h1.3 comes from this gene"
/codon_start=1
/protein_id="CAB03121.1"
/db_xref="GI:4008366"
/db_xref="SPTREMBL:P90884"
/translation="WKSTYVFLPYLLISFHVDAISMAAMSPWSSCTTCGGVS
ROLRCLTKSGSGEVRFRKVCQAKTCEKSRSLARDITCGGEIVSRGCEVYGRSLT
GANFLMYDDGTPCQAAATSRVCSKSCQIVGCDLISSEFRFDAGCGGGGCDN
GKFIWKEEETATACASNCDDIVDMAGRSASTSPICVYNATIGRVVPEKLCADK
LAPKVEARPCPMLTQPSRMADMTQEVPHCGEGRKREYICVOTAHVYTHVPTFC
ENGRTPAEBNCVSTSGRWEAGKSKTASGGQVRRRHVAACVGGSDCDEGRROE
TTTAGIPCSTATNSLDWNDRAVLDGNFSGSDNMQAPRLVAGESTCSTCGT
VMSRTVECVAVPISAPILKLPMSQCQDEQPKLFESCEVSRCPLOEDSKISEDAFY
OMRYGDMTCVASCISGKOKAKLKIOWTSKYSVMQSCDARRRPEKSRPCNOHP
PFWLTSKSDSCMSGCTARRSVCAQTVSKTDADAHYLRDRRCFFKPKQETETC
NVACPATWVSLSLKRHKIKLTKTKTQWTECSSCSGGERROWCEIIDSRSKTO
RPDVEDCANTRPQTVFVCSFSGCSRPBLLSNRVEFQNAEQKILIGIGVATLYOGT

SIKIKPARKFKKIKYMKNGKIKINDAHIKVANSGLRVEFHARMEDAGVCEFTDR
LOGNTVLNKYDFPASPVRDLKPOIPSTKROKOVSKEDVLEQASVILKMYVSL
IEALLTAPNDEKAREOLRKYGNELVARMIGHMSROKCTGVAGVQAQISCTYTH
GFRNVNDISICSLASVRPEPRPCRRPCRPMPASOMSECSOBCVSMIAOKRNV
PGRFNQTSVDIOHCDITNRRPATTNDPCRNQKAKRISDMSCSSECTGGVQURL
SCVWISSGRPACRNCBOMRRPHSARCVADEFLPCMPTASALYORDASQODSFRCD
IKLFHSCDLSVRQKCSCTCFYERKKR"
complement(join(18772..18944,19232..19286))
/gene="F53B6.9"
complement(join(18772..18944,19232..19286))
/gene="F53B6.9"
/note="CDNA EST EMBL:T02267 comes from this gene"
/codon_start=1
/protein_id="CAD30440.1"
/db_xref="GI:20338942"
/translation="MANOGRLPVFPVLLFAIVTAHSHSHSHSHHHHHGGYUG
GGYGVQRYGGGGYNGYNGYKROWYMKGRK"
complement(join(19892..20022,20241..20283))
/gene="F53B6.8"
complement(join(19892..20022,20241..20283))
/gene="F53B6.8"
/note="CDNA EST YK553f11.3 comes from this gene
CDNA EST YK553f11.5 comes from this gene"
/codon_start=1
/protein_id="CAB54255.1"
/db_xref="GI:5824510"
/db_xref="SPTREMBL:Q9U3E7"
/translation="MDFGILVFLMAVAVGTGAGISVSFSSHSHEPFRHLLGGCGFPNY
GGYQSNYMKGRK"
complement(join(20516..20602,20648..20953,20999..21424,
21609..21818))
/gene="F53B6.4"
complement(join(20516..20602,20648..20953,20999..21424,
21609..21818))
/gene="F53B6.4"
/note="predicted using GeneFinder
contains similarity to Pfam domain: PF00635 (MSP (Major
sperm protein) domain), Score=74.1, E-value=9.1e-19, N=1
CDNA EST YK193h3.3 comes from this gene
CDNA EST YK411f5.3 comes from this gene
CDNA EST YK193h3.5 comes from this gene
CDNA EST YK411f5.5 comes from this gene
CDNA EST EMBL:AW057173 comes from this gene
CDNA EST EMBL:AW057232 comes from this gene
CDNA EST EMBL:AW057344 comes from this gene
CDNA EST EMBL:AW057361 comes from this gene"
/codon_start=1
/protein_id="CAB03117.1"
/db_xref="GI:3877449"
/db_xref="SPTREMBL:Q9XVNI"
/translation="MIRSRGQPFRRSKCDQSDINRKNQFVLEKEARVCASIAFRML
VPLAIVTTSIPMYAIAIAFCANRRTYVAKRNNKSKSAKSKSTPGASIGKSR
SKAKHVSRSKSKKYGKSGKSGSKSKSKSKSKSKSKSKSKSKSKSKSKSKSKSK
TGVSDKQSNSSK
OVGSHLAEVNSIKHSKEMNAPAKLOQYQIGVNOELKMTSNERAYKIKGSDNSL
RYNVYGFAPRRSSVKIDVLRNCEKTDKVLITANANGSTNPHEAFANQAREHM
MVELVAA"
23555..24160
/gene="F53B6.5"
23555..24160
/gene="F53B6.5"
/note="Similarity to Rat Initiation factor 2 associated
glycoprotein (SW:12A6_RAT)"
/codon_start=1
/protein_id="CAB03119.1"
/db_xref="GI:3877451"
/db_xref="SPTREMBL:P90883"
/translation="MILSDQNFLOTQMKPEPTAOSKNTKESCEFGNSNEYKPGSLN
GQIAQORIRHEKIVPIYKGGEOARMENETIATFEGSGKGYFHDMDTSHYMKAF
ELABEKPIRLQKSKGLNLIDKRFATLAPRCRIDNIGETKYMALAKGCSYSTAAR
LMLTLTPELRQKAMVACILIOFAGNITIGRAVLLNLII"
join(26258..26342,26501..26802,26846..26926,26974..27107,
27153..27288,27334..27459)

gene
CDS

complement(join(17917..18025,18074..18331,18378..18555,
18604..18751,18800..18877,18947..19072))
/gene="R07E3.4"
complement(join(17917..18025,18074..18331,18378..18555,
18604..18751,18800..18877,18947..19072))
/gene="R07E3.4"
/note="contains similarity to Pfam domain: PF00153
(Mitochondrial carrier proteins), Score=236.0,
E-value=1.7e-68, N=1"
/codon_start=1
/protein_id="CAA89069.1"
/db_xref="GI:3878957"
/translation="MASTATIDLETSSKCLAGSAAAISKTTAPAPEDRVKLVLQ
LROSEFMAENGLRDCISKRLEOGAMALMRNGAGVARELNHILNFAERDLYNTL
LKVDNESEGFKIAGTEFVGSGGATLTFLMYPDPARTRLADYKKDGSKRYGM
DLCKTKIAEGVAWYKGLSALQPYSARAFPGFDPSIRSVEDPKSLNACRAI
AQITTSKGCVPIDTVRRSMMSGOKIOKTSTKDCKTLVKDKGNFYRALT
NSLRSGALLTFYEFSKYM"

gene
CDS

complement(join(21213..21359,21654..21739,21787..21886,
22011..22170,22220..22335,22428..22542,22637..22779,
23254..23259))
/gene="R07E3.3"
complement(join(21213..21359,21654..21739,21787..21886,
22011..22170,22220..22335,22428..22542,22637..22779,
23254..23259))
/gene="R07E3.3"
/note="similar to cuticulin"
/codon_start=1
/protein_id="CAA89068.1"
/db_xref="GI:3878956"
/translation="METSPRGFLSTNNVVAENPELTKNDPKEVQCFYMERRIO
KYIOLSPMPHKSOLAMPVCKEYLDSSPGPRPYRAYTGOMYHHMTCDTERHT
FCMLHSCPVDDNGCRVGLDMKGCALDKLTLLILEFTDMAERGEAHVRKADRDN
MYFDQISTVEPGIDCVDPSCPDRRRSRNTLPADNITAIAHIEYDESI
LSYIIIPNDIISLMQRNFMDRISELCTAIGTLVFLNAFFIISLVIVHCDF
RTSPKLEKRTSLM"

gene
CDS

join(23363..23562,23642..23720,23765..23899,23955..24269)
/gene="R07E3.2"
join(23363..23562,23642..23720,23765..23899,23955..24269)
/gene="R07E3.2"
/codon_start=1
/protein_id="CAA89074.1"
/db_xref="GI:3878962"
/translation="WLINNSGPLPLTSLYLIPDLLNNVRFEMIDRPACTREPAL
TDYISRADKPMSIHLLIFESQCVLLESFQLASLISVTGEVVDTAEAVIMEDI
EVADVETVAIVMEDVAEVTEEAAMEAVEETEEVTAADMAVFADTEEVIVM
EATTEMEEAIVMEDVAEVTEEAAMEAVEETEEVTAADMAVFADTEEVIVM
TEVARKEGHOCOTKOCRLDLS"

gene
CDS

complement(join(26570..26749,26841..26917,27073..27251,
27330..27815,27863..27958,28006..28071,28119..28220,
28273..28285))
/gene="R07E3.1"
complement(join(26570..26749,26841..26917,27073..27251,
27330..27815,27863..27958,28006..28071,28119..28220,
28273..28285))
/gene="R07E3.1"
/note="contains similarity to Pfam domain: PF00112 (Papain
family cysteine protease), Score=215.1, E-value=3.5e-61,
N=1
CDNA EST EMBL:M89104 comes from this gene
CDNA EST YK68h10.5 comes from this gene
CDNA EST YK72g11.3 comes from this gene

Query Match 5.8%; Score 67; DB 3; Length 39736;
Best Local Similarity 64.5%; Pred. No. 3,5e-05;
Matches 100; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

OY 641 GATGTCCTTGTAAATATTGGTAATAAATTGGATATCCACAGTTTAAATGCCGC 700
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 22171 GAAGCGCCTTCGCACAATATCCTTCACAAAATCTTGAGTAGTCCACAGCCTGATGCGG 22112
OY 701 GCCAAGAACCTCACGTATCAAAATATGCGCATGCATCACGCTTTTCATCAATGCCAGA 760
|- - - - - ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 22111 GAAGGAAGACCACGATATCAAAATATGCCGATAGATATACATGATATTGTATGTGC AAA 22052

OY 761 TCAGTATTACCATTTAAAGACCAAAATAGCGAATGT 795
|- - - - - ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 22051 TTTGATACACGTCAAAGAGCCTGGCTTACTAT 22017

RESULT 24

CENM142 36306 bp DNA linear INV 24-JAN-20020

LOCUS Caenorhabditis elegans cosmid M142, complete sequence.
DEFINITION
ACCESSION Z73428
VERSION Z73428.1 GI:1657678
KEYWORDS HTG; Cuticulin; Homeobox protein EVX1 like.
SOURCE Caenorhabditis elegans.
ORGANISM Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida;
Rhabditoidea; Rhabditidae; Pelodierinae; Caenorhabditis.

REFERENCE
AUTHORS 1 none.
TITLE Genome sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium
JOURNAL Science 282 (5396), 2012-2018 (1998)
MEDLINE 99069613
PUBMED 9851916

REMARK The C.elegans Sequencing Consortium.
REFERENCE 2 (bases 1 to 36306)

AUTHORS McMurtry,A.A.
TITLE Direct Submission
JOURNAL Submitted (21-MAY-1996) Nematode Sequencing Project, Sanger Institute, Hinxton, Cambridge CB10 1SA, England and Department of Genetics, Washington University, St. Louis, MO 63110, USA. E-mail: jes@esanger.ac.uk or rwenemate@c.wustl.edu
On Nov 4, 1996 this sequence version replaced gi:1370039.
Coding sequences below are predicted from computer analysis, using predictions from GeneFinder (P. Green, U. Washington), and other available information.
Current sequence finishing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note.
This sequence is the entire insert of clone M142.
The true right end of clone C44B9 is at 5059 in this sequence. The start of this sequence (1..104) overlaps with the end of sequence Z73424.

COMMENT

The end of this sequence (36203..36306) overlaps with the start of sequence 299276.

For a graphical representation of this sequence and its analysis see: <http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=M142>

IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

FEATURES

source	Location/Qualifiers
gene	1..36306
CDS	/organism="Caenorhabditis elegans" /db_xref="taxon:6239" /chromosome="III" /clone="M142" join(3511..3570,4639..4750,4821..5201,5689..5795) /gene="M142.1" join(3511..3570,4639..4750,4821..5201,5689..5795) /note="cDNA EST yk258a1.3 comes from this gene cDNA EST yk532a11.3 comes from this gene cDNA EST yk532a11.5 comes from this gene" /codon_start=1 /protein_id="CA97807.1" /db_xref="GI:3878736" /translation="MRPIYDLSITSPFLSLILICSNPINDGLVDSLEIHCVTHK AVEVILLDSGSGIDTTEKQSLFAMHLASRLNISGDSHMLIOYAEPRKLEFSG QFNHPQLEMAIORIEFGSATNGOALRLTLKGLGQARGIPKVAIVITDGSODD VSEPSOLIRADPMYAIAGVTNLVNHQLOHMTGNPVRYFTVESFQDLRALDSLTM SKCTERPRETPETITGPRITIGYKASTKQPEGNVFMVMDTHDEECAGKEFPDSKS IGLVPSACNVHRYSLNPKGIFVEASTVFMHSLFMTDTQTVKQCEKMDKRV TVPLSVMTITVREQIYQMPQCAIVLRKAGPGRIVREFATLCESYVHMECELEVA ADKDFGMLVHSCVYDNGGIDRVLDISNGGIDAVLSTPDYLSLRLAKRPVHVEY ADRPVFOFQOCLLKVDGCEGITPPQCKKLGPDGHHNHHPKRRKRLVRRLAD GVGTIDVTSVTVLEQEPACQVTTSEPRSTSTNCKLENS" join(17427..17468,18628..18766,20758..20915,22487..22677, 22840..23053) /gene="M142.4" join(17427..17468,18628..18766,20758..20915,22487..22677, 22840..23053) /note="VAB-7 even-skipped homologue, contains similarity to Pfam domain: PF00046 (Homeobox domain), Score=111.1, E-value=7.1e-30, N=1" /codon_start=1 /protein_id="CA97809.1" /db_xref="GI:3878738" /db_xref="SWISS-PROT:Q93899" /translation="MOSFIDSLIGVAKVPSLVEMVAASRASFSPPEQOHDPMGV VAAAAAAGRHHPYDNRDQGMRRRTAFSREOIGRLREFAKENYVSRKTRGEIA AEINLPEGITIKVFNQRMKDKQVRVGLAMPPEPPQMAAYMLNPFVEMMKTAASQ
gene	join(8540..8572,9335..9416,9468..9679,10373..10721, 10772..10974,11033..11175,11871..12501) /gene="M142.2" join(8540..8572,9335..9416,9468..9679,10373..10721, 10772..10974,11033..11175,11871..12501) /note="predicted using GeneFinder contains similarity to Pfam domain: PF00092 (von Willebrand factor type A domain), Score=155.0, E-value=4.3e-43, N=1 cDNA EST yk518d1.5 comes from this gene" /codon_start=1 /protein_id="CA97806.1" /db_xref="GI:3878735" /db_xref="SPTREMBL:021540" /translation="MRPIYDLSITSPFLSLILICSNPINDGLVDSLEIHCVTHK AVEVILLDSGSGIDTTEKQSLFAMHLASRLNISGDSHMLIOYAEPRKLEFSG QFNHPQLEMAIORIEFGSATNGOALRLTLKGLGQARGIPKVAIVITDGSODD VSEPSOLIRADPMYAIAGVTNLVNHQLOHMTGNPVRYFTVESFQDLRALDSLTM SKCTERPRETPETITGPRITIGYKASTKQPEGNVFMVMDTHDEECAGKEFPDSKS IGLVPSACNVHRYSLNPKGIFVEASTVFMHSLFMTDTQTVKQCEKMDKRV TVPLSVMTITVREQIYQMPQCAIVLRKAGPGRIVREFATLCESYVHMECELEVA ADKDFGMLVHSCVYDNGGIDRVLDISNGGIDAVLSTPDYLSLRLAKRPVHVEY ADRPVFOFQOCLLKVDGCEGITPPQCKKLGPDGHHNHHPKRRKRLVRRLAD GVGTIDVTSVTVLEQEPACQVTTSEPRSTSTNCKLENS" join(17427..17468,18628..18766,20758..20915,22487..22677, 22840..23053) /gene="M142.4" join(17427..17468,18628..18766,20758..20915,22487..22677, 22840..23053) /note="VAB-7 even-skipped homologue, contains similarity to Pfam domain: PF00046 (Homeobox domain), Score=111.1, E-value=7.1e-30, N=1" /codon_start=1 /protein_id="CA97809.1" /db_xref="GI:3878738" /db_xref="SWISS-PROT:Q93899" /translation="MOSFIDSLIGVAKVPSLVEMVAASRASFSPPEQOHDPMGV VAAAAAAGRHHPYDNRDQGMRRRTAFSREOIGRLREFAKENYVSRKTRGEIA AEINLPEGITIKVFNQRMKDKQVRVGLAMPPEPPQMAAYMLNPFVEMMKTAASQ

BASE COUNT	ORIGIN	Query Match
11616 a	6626 c 6322 g 11742 t	5.4%; Score 62.6; DB 3; Length 36306;
gene	FGATPGNAGVNGNNGSTSPSAGSLPFLPPLGPPSLQNSTKSPSSPHSDSSKSK MTSSDDDESKPVNFNSRNPSSSSSPSYTDP" join(28250..28366,28422..28499,28558..28991,30672..30966, 31500..31763,32933..33043) /gene="M142.5" join(28250..28366,28422..28499,28558..28991,30672..30966, 31500..31763,32933..33043) /note="predicted using GeneFinder cDNA EST EMBL:T01774 comes from this gene cDNA EST yk470a9.3 comes from this gene cDNA EST yk470a9.5 comes from this gene cDNA EST yk476b6.5 comes from this gene cDNA EST yk563c12.3 comes from this gene cDNA EST yk662c9.3 comes from this gene cDNA EST yk679e8.3 comes from this gene cDNA EST yk563c12.5 comes from this gene cDNA EST yk584c4.5 comes from this gene cDNA EST yk662c9.5 comes from this gene cDNA EST yk679e8.5 comes from this gene" /codon_start=1 /protein_id="CA97808.1" /db_xref="GI:3878737" /db_xref="SPTREMBL:021541" /translation="MASCNTASVATGACVEIETNGLSARGVTSFDTTRKVLVD TKEMAIKPKPIRIFNSHLKNTIVLSALDESKPRAKTEQFQANNPVGCRTTERT DKTGLGELPLPKSPATISIGQAVYLDLKTPTDTCGCGDILVGLVYHKKREYSD VTDAKATGFDESRAESAVLQOVKILSKPTEYTAARPPLDTFGALVANSSESMDST DDIIPOLASDTLALSMFOAQKIDLOLQGIIEKIDEDLOLQFWDDETSTRVLAE GVAALGESBARARIGCVSPPLVFEHFEHTEKQIDLOLQFWDDETSTRVLAE VHYDKRHPDLPAELAKFVVIADPFLCAEIKASHIRLLKSDVVLCTGTAI MEDYASRLMAHHTSPFEPHANNLANDESCFANVORTTFC" join(34039..34101,34448..34750,35530..36306, 299276.1:1:105..698,299276.1:1:1364..1566,299276.1:1:1666..1772, 299276.1:1:2495..2795,299276.1:3:714..3928, 299276.1:4:056..4104,299276.1:5:846..5958, 299276.1:6:478..6729,299276.1:7:169..7338) /gene="M142.6" join(34039..34101,34448..34750,35530..36306, 299276.1:1:105..698,299276.1:1:1364..1566,299276.1:1:1666..1772, 299276.1:1:2495..2795,299276.1:3:714..3928, 299276.1:4:056..4104,299276.1:5:846..5958, 299276.1:6:478..6729,299276.1:7:169..7338) /note="similar to Zinc finger, C3HC4 type (RING finger) cDNA EST yk49h8.5 comes from this gene" /codon_start=1 /protein_id="CA97810.1" /db_xref="GI:3878739" /db_xref="SPTREMBL:045962" /translation="MAPTGGGCGGQVLECCSICNHFNEFLFVLSLIGHYTRKCAE KPNQNTKPCDHMDKMTTHSPSEYPMNVALLSVIFPRKQCTLGSAGVSEAKRDQSLI QIAKFEREADSEKGVTSREIERTLDKRLVLAICQVREVDRLKTLKRCISEV MIELIISQNTFVSQLSMAGVARGQFGLAPMODVRLILMLTLETGICIRKINLV MYVQVITASQIPQVSKTCVHVVOLLYADQSCFVNVLKDGESSLMQKEEPRYSIR EHSOIVQIAFESGLRIGPQWGAALLYADQSCFVNVLKDGESSLMQKEEPRYSIR LASQTSMLYPAIRYPLTYQVIRPLEFAGIEHEDTSRMIGDALHQIRILILKHCSD DLKMKPEERKGVYILQAEVPGMGGGGGGAGAGTIGLHLPIISQIDTGRSISIT NPKDNSHNSPQTPPKQPRKRYQMGIPRPMGSSSDAPPIPSHQOQPPQFNSQHL PQREFGGRGAGAPPPEPPQPMMLIGDMPGAPMOQTEVLTDAGQVNTPOPVVIM QSPHLHGGGVWMIPOQVMPPOSMTVGGPMPGMPKPSIIPVQVPTMTATASP TGSIVYPAASPGQPPHTWIOSIGVKKRSKRLTISFLNFYDFPLIRKLKK EKKGADIDEEFEKIKSTDFKTPSSFSSTQDNMPKIDRGSGAGYWGFGTYLRSGADA EQLLAKYETILKRLQSDDEDDPDDGIGHVSTVASVLDLDRHDPPLTWIPVPTD LPAIPISFANMPTEETMTMIGEMVQNPRAPSLTVASVLPNPNVNASASATVQAEEN RKILIDFLKRYKMTLMEKYSTCFHYTLKDYVLFYVLANLNASVPRRRRAATTIPQ PVLPVQVPOVQVIVPAENPNPNVPPVPPPOCPMLVDSAGILPPIRLIVAAHON VNSLSLDKIVDKERTISEQGNASSEANMLRLRELIRRAESQMAHIDPYTNNCLLAL QOYDMEIQOHLHLPYEG"	11616 a 6626 c 6322 g 11742 t

FEATURES

CDS

gene

CDS

gene

CDNA EST yk679g8.3 comes from this gene; CDNA EST yk239h1.3 comes from this gene; CDNA EST yk281d10.3 comes from this gene; CDNA EST yk302e2.3 comes from this gene; CDNA EST yk303e9.3 comes from this gene; CDNA EST yk339h1.3 comes from this gene; CDNA EST yk383d9.3 comes from this gene; CDNA EST yk417g8.3 comes from this gene; CDNA EST yk426d4.3 comes from this gene; CDNA EST yk477f1.3 comes from this gene; CDNA EST yk482d12.3 comes from this gene; CDNA EST yk525f10.5 comes from this gene; CDNA EST yk540b10.5 comes from this gene; CDNA EST yk580a3.3 comes from this gene; CDNA EST yk580a3.5 comes from this gene; CDNA EST yk612a9.5 comes from this gene; CDNA EST yk612e9.5 comes from this gene; CDNA EST yk615c12.3 comes from this gene; CDNA EST yk643h10.3 comes from this gene; CDNA EST yk643b10.5 comes from this gene; CDNA EST yk670g10.5 comes from this gene; CDNA EST yk679g8.5 comes from this gene

/codon_start=1
/protein_id="CAA95854.1"
/db_xref="GI:3880400"
/translation="MTDVPQSPVEDDQAQOASYYDVLGNAYVQFAINAYTKTEPH
PLINSTNGAEKRSYVGVNAYAAKAYDGTNSYYVPRKNAYVSGTERRATVAESG
KCAIYVGTGFGAAYVLTQESLASAGGAAVLEVDQSAKSLSTIREALAY
EHRTFSALHQQRILAMVPEKITEENTSLDLTLDVAQGLAIETPPSVNLITGQRAV
NLASIVLQESKRAHDHYDIPNERANRLEQSLQSVFLDLIVREKQVTEKSNELS
TSVFPFKLTIEEAQKQYKVAPEEMLMKH.IQSTSEQLSTDLQSLREKGVNFECDYKID
STIDYLENKRPNTDAEYDQYKRVDEVLNQRRIEQLSTWTLSTLIIISAEQFEPEDL
LLEELFYDPQPVPTIRNLYKRNRA"

gene
101n118064.18085.18138.18325.18375.18520.18908.19295
13346.19445.19506.19721)
/gene="W01A8.3"
/note="Similarity to C.elegans cuticulin (SW:0071_CAEEL)".
/codon_start=1
/protein_id="CAA95850.2"
/db_xref="GI:15718222"
/translation="MRIMRGSTPKVICAENDLADIVTSKPPRCNIEFVKGRADKSC
RSYVANGNSYSLPLGKGMORLSANPDRGVNCEVAVYVSHPGFTLINDRAFHK
CTYMPDELVIONIVSMPTPELSDSNVPMGSESVRRDGNPGFTLYANNGDITVH
VMECTPADMGHLYKKCFYTDGDEGHAYVDPGCAITDPLISELSDLSMRAHASSO
VKVADSLVLTFTQIRLCQKQMGKQCEYTPNCCVKKLILSASDGNKTRKADRS
YEIDVATSNLVLTPADRLGLAPSPFCVPRLLPLPLITIVSLTVASVALVTRQ
NHKKELDIMOSGRYY"

gene
complement(20058.20321)
/gene="W01A8.4"
complement(20058.20321)
/gene="W01A8.4"
/note="CDNA EST yk100d3.3 comes from this gene
CDNA EST yk100d3.5 comes from this gene
CDNA EST yk607d1.3 comes from this gene
CDNA EST yk607d1.5 comes from this gene"
/codon_start=1
/protein_id="CAA95851.1"
/db_xref="GI:3880397"
/translation="MDPAMFRVYSADMTQAEFRFPTPTVLYGVGFALFYITRLM
FVPMKSNBACIDKGLTMMWDRQRTLTQSGGLAPRTIGIDF"
join(21264.21365.21414.21568.21618.21843.21891.22068
22134.22408)
/gene="W01A8.5"
join(21264.21365.21414.21568.21618.21843.21891.22068
22134.22408)
/gene="W01A8.5"

gene
join(21264.21365.21414.21568.21618.21843.21891.22068
22134.22408)
/gene="W01A8.5"

CDN

JOURNAL Submitted (27-SEP-2001) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA

REFERENCE 6 (bases 1 to 28396)

AUTHORS Waterston R.

TITLE Direct Submission

JOURNAL Submitted (22-MAY-2002) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA

COMMENT

Genome Sequencing Center
Department of Genetics, Washington University
St. Louis, MO 63110, USA, and
Sanger Centre, Hinxton Hall
Cambridge CB10 1RO, England
email: rwaterston@wustl.edu and jessesanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one ml3 subclone.

For a graphical representation of this cosmid sequence and its analysis see:
<http://www.wormbase.org/db/seq/sequence?name=F10E7;class=Sequence>

NEIGHBORING COSMID INFORMATION

The 5' cosmid is T02G5, 200 bp overlap; the 3' cosmid is T07F8, 1000 bp overlap. Actual start of this cosmid is at base position 1 of F10E7; actual end is at 7817 of T07F8.

NOTES:

Coding sequences below are the result of integration and manual review of the following data: computer analysis using the program GeneFINDER (P. Green and L. Hillier, personal communication), the large scale EST projects of Yuji Kohara (http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html) and The C. elegans ORFome cloning project (<http://worldb.dfci.harvard.edu/>), similarity to other proteins from Blastx analyses (<http://blast.wustl.edu/>), sequence conservation with C. briggsae using Jim Kent's WABA alignment program (Genome Research 10:1115-1125, 2000), individual C. elegans Genbank submissions, and personal communications with C. elegans researchers. TRNAS are predicted using the program tRNAscan-SE (Lowe, T.M. and Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).

FEATURES

source

1. 28396
Location/Qualifiers

/organism="Caenorhabditis elegans"

/strain="Bristol N2"

/db_xref="taxon:6239"

/chromosome="I1"

/clone="F10E7"

/complement(900..4554)

/gene="F10E7.8"

/note="for a graphical representation of this gene see: <http://www.wormbase.org/db/seq/sequence?name=F10E7.8;class=Sequence>"

CDS

complement(join(900..1126,1176..1637,1685..1769,1985..3460,3510..3689,3736..3885,4119..4291,4518..4554))

/gene="F10E7.8"

/note="coded for by the following C. elegans cDNAs:
yk44g1.3, yk20f8.3, yk12b7.5, yk16g12.3,
yk16g12.5, yk20f8.5, yk36g6.3, yk36g6.5, yk44g1.5,
yk36h11.5"

gene

CDS

/codon_start=1
/product="Hypothetical protein F10E7.8"
/protein_id="AA082421.1"
/db_xref="GI:1086830"
/translation="MPAHLNKAECRIYGDKERRSADDEHYEYSSSLAARRSANVTK
LKQLEAPMTIPATGFTIHSSDNDKHDFEFTFEESVQKNVYIYDIDCSMES
ADLEVRKLSARITLFLQGSADDFCLMEIEKEEYDQKRLMNPVIRTKRMDPEGS
MYVHAIKNSFICVKNQGYOALCTLLMEIEKEEYDQKRLMNPVIRTKRMDPEGS
LSDSEKWRTPPTVNDNEMRLVILAGIYHIMVOLDRESREDPTDRLKAFRE
EYDEPIENGNEPLILFDMLHPHYGTAPHPRIKILITLILGLGMEITVR
KREKMTAGDIMEPTITVAASQAFKADDEHRRNMAHRRGSPARGMAMARQAYND
DSKDDAYSSEASDSTASKKEDSPGSSSTPSIPDTASNFRQSGEQPRGSPV
FPVQKTLTPMKSKTNMADIEDIQRGRKFNDFDGDPTLFLGPPSCGAVKIL
RNMTYSTLEIQVQKDEIRLNYLSMKEDIPETKTELRYRKLIPNLISQICALAKVM
SSVPTKARHGLANVLDCLPMEASDILNSISLDSSTSPSEGRALIDNRK
ELVKSALITLLVYKHRLNHIYOFEYICQOIVYVNIPLIKFLDONTKLIOSR
ELVAYNPQCILYHYRNNENPILITODNIEPRPREGPYFMWRVFPAINLILNK
LVKANDVAKMLMFKSAPVLRLEKVSAYLQVLVLAIMQSRKLRKRSNMDI
ISATSRVRRHRTDWFAPASDIKRCDYQKDSLIKASIERFHSRRSALYPOAIFRE
NDAPWPGDGYLNRYDMRDFEVDICAHSLGANLKLGRHFKDYEKWLEQVFNASID
MDKLIERGVEDLM"
4936..5624
/gene="F10E7.8"
/note="for a graphical representation of this gene see: <http://www.wormbase.org/db/seq/sequence?name=F10E7.7;class=Sequence>"
join(4936..4985,5210..5372,5417..5554,5601..5624)
/gene="F10E7.9"
/note="contains similarity to Pfam domain: PF01247
(Ribosomal_L35he), Score=256.8, E-value=9.6e-74, N=1;
coded for by the following C. elegans cDNAs: yk51f3.5,
yk64g10.3, yk99d1.3, yk99d1.5, yk115e7.3, yk115e7.5,
yk197h7.5, yk207c1.3, yk207c1.5, yk221f10.3, yk229g1.3,
yk229g1.5, yk282e4.3, yk282e4.5, yk298e8.5, yk308e6.5,
yk402a7.3, yk438e5.3, yk449b9.3, yk449b9.5,
yk450d5.3, yk450d5.5, yk460t8.5, yk469g1.3, yk468d11.3,
yk468d11.5, yk480c10.3, yk480c10.5, yk482f6.3, yk482f6.5,
yk64g10.5"
/codon_start=1
/product="C. elegans RPL-33 protein (corresponding
sequence F10E7.7)"
/protein_id="AA082422.1"
/db_xref="GI:1086831"
/translation="MADPVAARRSAPPTGRLVYKAIFTGKRLRQSEHTSLKLE
GVNKKEDFAGKRRVYLYKAHKKTLTGHTVTFRAIWGKITRBNAGAVARAF
HHNIPDSGLGRIVLYPSNI"
complement(5858..9377)
/gene="F10E7.9"
/note="for a graphical representation of this gene see: <http://www.wormbase.org/db/seq/sequence?name=F10E7.9;class=Sequence>"
complement(join(5858..5968,6012..6221,6265..6633,6769..7271,7551..7752,7801..8020,9268..9377))
/gene="F10E7.9"
/note="coded for by the following C. elegans cDNAs:
cm13g1"
/codon_start=1
/product="Hypothetical protein F10E7.9"
/protein_id="AA082423.1"
/db_xref="GI:1086832"
/translation="MSETGSSSHSVVEPPDSIFSGEETADSSQLPMWORTITLTKDK
VLFGTGVGFVAVWNIIFGLIFRLMIYGTAVANVILGLCTSLALITVYSAIG
IVERCOIKSGGIYFLVSHVIGHOIGAMIGIYAGOVANVILGAGSEVANHFESE
SKMIKTAITLITVAVLTAVNAGTYTVVRQDIYLLITLALAVDFIFGALPSSPEBEG
VFRSSSEKIRYNASHYEAVNCSITIGIPROIPEOSFTVFGVFANLGVLAGVNMKG
DLKDPKHSIRPLGELSAVGSSTICFIFIMIGDGMFLCDPMANLGVLAGVNMKG
VGLVYCSLSIVGNDNPVLAVGLMCAVSFLVGLDNLDAIISTPRLITVAYVY
SYVLSAYSLATQVHADDEVETVGTGNKNKNDNELEPBEROVNITSTETETGIS
QEAATVWSEMSNRVYSLGFLVNLAILLILHPFVAVFVAILAYWYIGVCPGCGG
ISTFSLCHMRKTVYSSLDLSRGATTTLPETTFISGVSSSQOENANDYDROROYHNS
ENVSQHLD"
11359..12212
/gene="F10E7.5"

CDS
 /note="for a graphical representation of this gene see: <http://www.wormbase.org/db/seq/sequence?name=F10E7.5;class=Sequence>"
 join(11359, .11386,11435, .11493,11541, .11618,11668, .11846,
 11894, .12212)
 /gene="F10E7.5"
 /note="similar to 60S acidic ribosomal protein Po (L10);
 coded for by the following C. elegans cDNAs: CEES189F,
 CEESR40F, CEESN26F"
 /codon_start=1
 /product="Hypothetical protein F10E7.5"
 /protein_id="AA82424.1"
 /db_xref="GI:1086833"
 /translation="MARSRPRKNVSLTKVKKTKDTKNTLNVEYRASVDYQKNLFIET
 IANNRSTRFLAIRKRYKENSRFFPGKNVNVISIALGKSDPYANQLHKASILTKGCCG
 LMTFMKSRKEAEAFSEASEDIAVDVATVPLBGPISQFAFSENPOLRLGLP
 TKLDKGVITLYIQEFVECKEPELVEQAKILKHFVEKMSQFLIFKAHWNKKDKFEL
 DA"
 12451.12959
 /gene="F10E7.6"
 /note="for a graphical representation of this gene see:
 Query Match 4.7%, Score 54.8; DB 3; Length 28396;
 Best Local Similarity 52.4%; Pred. No. 0.025;
 Matches 175; Conservative 0; Mismatches 147; Indels 12; Gaps 2;
 QY 448 ATGCAGATAGCCGTTATGAATAATTTGGATGSGTGGACCAACCGGTCACACTTCAATT 507
 Db 14488 ATGCCAAGTGAATTAATTTTCGATTAAGAAACAGGAAAAAGATGTCCACCGATACATAT 14429
 QY 508 GCTATCATTTGGTCACGCCAGTTTATCATTAATGAGCATGCGATTCTGAAACCGTTATACT 567
 Db 14428 GCTTCATTTGGAGACAGATGATTCATTCATGTTTGGATTTGAT-----GGAATCAT 14378
 QY 568 TTTCGGCGGGTTTCCATATTCCTGCTTTGTCCATGATGTTACGGTGATACGTGAAATT 627
 Db 14377 AACGGGATTCTAGTACAAATTTGCCATGTGGAGATGACAGGAAATTAAGATTCTTATT 14318
 QY 628 CTAAATGCTGATGATGATGCTCTTGTATTAATTTGCTAAATTAATTTGGAATATCCACA 687
 Db 14317 ATTTGATCAAAATGATGTGGAATGATGATCATTAATGATGATGATGATGATGATGATG 14258
 QY 688 GATTTA---ATGGCTGGCCAAAGACGCTACGATACAAATATGCGGATGATCACAGCTT 744
 Db 14257 GAACAATGATGTGCAATTTCAAGAAACATCATGTGTTCTCAAAAGACGTAACA 14198
 QY 745 TTCTATCATGTCAGATCAGATTATTCATTAAAG 778
 Db 14197 CGTTTATCTGTCAATCAAGATTGTGTATGAAG 14164
 RESULT 29
 CEM28 40265 bp DNA linear INV 13-AUG-2002
 LOCUS Caenorhabditis elegans cosmid M28, complete sequence.
 DEFINITION
 ACCESSION Z49911.1 GI:872077
 VERSION 249911.1
 KEYWORDS HTG; Caselin kinase I; Cuticulin; Esterase; L7AE; Ribosomal protein;
 SH3 domain; Transposon.
 SOURCE Caenorhabditis elegans.
 ORGANISM Caenorhabditis elegans.
 Eukaryote; Metazoa; Nematoda; Chromadorea; Rhabditiida;
 Rhabditiodea; Rhabditiidae; Pelodierinae; Caenorhabditis.
 REFERENCE
 AUTHORS none.
 TITLE Genome sequence of the nematode C. elegans: a platform for
 JOURNAL investigating biology. The C. elegans Sequencing Consortium
 MEDLINE Science 282 (5396), 2012-2018 (1998)
 MEDLINE 99069613
 REMARK The C.elegans Sequencing Consortium.
 REFERENCE 2 (bases 1 to 40265)
 AUTHORS Gardner,A.E.
 TITLE Direct Submission

JOURNAL

Submitted (23-JUN-1995) Nematode Sequencing Project, Sanger Institute, Hinxton, Cambridge CB10 1SA, England and Department of Genetics, Washington University, St. Louis, MO 63110, USA. E-mail: jes@sanger.ac.uk or tw@nematode.wustl.edu
Coding sequences below are predicted from computer analysis, using predictions from GeneFinder (P. Green, U. Washington), and other available information.
Current sequence finishing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note.
IMPORTANT: This sequence is not the entire insert of clone M28. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.
The true left end of clone M28 is at 1 in this sequence. The true left end of clone D2089 is at 40166 in this sequence. The right end of clone C14M4 is at 5382 in this sequence. The start of this sequence (1..101) overlaps with the end of sequence Z49909. The end of this sequence (40166..40265) overlaps with the start of sequence Z36948.
For a graphical representation of this sequence and its analysis see: <http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=M28>
IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

FEATURES	SOURCE
gene	CDS
gene	CDS
gene	CDS

```
Location/Qualifiers
1..40265
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="II"
/clone="M28"
join(249909,1)..38635..38695,Z49909,1..38912..39184,
Z49909,1..39393..39575,Z49909,1..39631..39784,102..242,
292..412,462..924)
/gene="C14M4.13"
join(249909,1)..38615..38695,Z49909,1..38912..39184,
Z49909,1..39393..39575,Z49909,1..39631..39784,102..242,
292..412,462..924)
/gene="C14M4.13"
/note="contains similarity to Pfam domain: PF00066
(Eukaryotic protein kinase domain), Score=36.9,
E-value=1.2e-09, N=1"
/codon_start=1
/protein_id="CAA90134.1"
/db_xref="GI:3678698"
/tranlation="MAAVVAVKQKEGCKQLSTIINGFNAVOMIGAGVYEV
RRSPMTPRACKAEALIDHNINLKTMDMLTLKDNRSHNITGVLSERFNIVH
LVSPSLSDIKRTAPVKNTFTLTFTTAVCAICDPSISIORITGYELRDVKSFAIVG
SEBEKYVYVDLGECRNPNKKELRKRRKARPFGTILYSLNINHORMEGRDDPW
SLTNMIETHSDLPEWNESENEDTKKAETKIDGLGAPEEFMRIRYLLTYSE
PDVXKIREVLQILMTSKFKTPDMPLDWKGCCAEIFKPQAIVAHGKTRKVSIAL
LDLPKRGVTERDEFPQGEQEPMRDSVSDDTIIEKGGPPPAKPVSKPPRP
GYMSKLPOLAOORCSAEBHKDLSTPPAIARSPDIGQTTMKDPVAIAIPPIEKHSV
AAVNAIPSRSKSNRKPAPH"
join(2766..2816,3520..3577,3647..4214,4456..4603,
4867..5009,5322..5712,5991..6251,6448..6547,6672..6895
)/gene="M28.1"
join(2766..2816,3520..3577,3647..4214,4456..4603,
4867..5009,5322..5712,5991..6251,6448..6547,6672..6895
)/gene="M28.1"
/note="similar to cuticulin 1
CDNA EST EMBL:T00366 comes from this gene
CDNA EST YK389e11.5 comes from this gene
CDNA EST YK510a3.3 comes from this gene
CDNA EST YK512f1.3 comes from this gene
CDNA EST YK647d10.3 comes from this gene
CDNA EST YK659c11.3 comes from this gene
CDNA EST YK664a5.3 comes from this gene
```

```

cDNA EST yk510a3.5 comes from this gene
cDNA EST yk512f1.5 comes from this gene
cDNA EST yk647d10.5 comes from this gene
cDNA EST yk659c11.5 comes from this gene
cDNA EST yk664a5.5 comes from this gene
/codon_start=1
/protein_id="CA90131.2"
/db_xref="GI:5824557"
/translation="MTISIVLPICVLLILGAPKTSAGSASAFDVAITGPIYISQ
RDRVIVTEIRLFTGKVFYKGDYANPQCTKTFENGVASSGGHLEDFTFLGYTKRE
PIYEDHLKASRGSSDSKI5NOYONKATSMNENHASKSGIDLTGKSTKTHENDINSA
PDLAEFLPRKNETTSDTEDEGKWKIGGASAGNAIFGTRGELRFGKSTYSSRT
RSRPGSTPLPSTSSPDCPLTTCAPCAECERRRATNGIRLEVPINSCNTRODRT
NPSVSVSLIAVVSFHSFTIKLDKAYHOCVCAAEAKYSTDLDVMTDEQELNGTV
EPSCDYLISDONGSNYSNIVGEVLAHVHOCVGGGLTKMLHOCYVAGAGQOE
VIOHGCTLDQMLQTPYTSDEGSAVDVYIFKFPDRSTVDRCTTTFCSVDADCL
DMTPKCGANNLIRKRRSHSLSSLSHANSITLVEDIDSSKTDITMOLPTALLRK
TFEEYERTFCVSVASFGLISASTFEATISIAVOFLVSFCFOKTEKLAERSTNPL
SMRYSEIRKILYFKLISSPVEFONAVFEVYFCLYTKIOSKQNCQSVFELFN
IPDVFNSIRKKNVILICLAL"
join(8730..8930,9116..9304,9425..9581,9781..9941,
10191..10348,10393..10476,10521..10679,10729..10852,
10901..11120,11492..11672,11713..11895,11940..12020,
12065..12289,12604..12677,12964..13056,13233..13358,
13452..13530)
/gene="M28.8"
/protein_id="CA90132.1"
/db_xref="GI:3878696"
/translation="MNLCSVFGISICILASSFAQTSPOCNOTTPOIKSMKPNFGASN
AFYDIAGKISDGLONTPNKSTYDIDLDFGNGAVSYNOVDVINOIGALITPNSVG
FVFLAALFVGATCIWICFCSCVPKSTRESEITVYAGLILVFNATLGLITFN
VAETDLTSDINGMYIINOTASDINVTINDGDOVTCETTTTGTTCGVANIPISYIS
AIYVGCTATGELTAVNAFEYRNFOINATIDANILITSDMNTOTFGVNIPIYIS
LQOQIEIMIPTLALSSGAQIVTSPETIIALNDVIDLKRNIOTQDASGSINSTOG
TINDMSTIRNLGNVQNDLNNMNSLASARHDVYSSSGIKIGIRLLITIPACIG
CLCEILAEVAVVLSLKKEVEFLATVAVYVILISIVLIFSSAEVAVGLTSA
ACVPIFEDNNVTFLRLMNOTIAPVGTGSPDYVIGILOSQNPMTLYTAVDGOFTI
SADSINOONILASRYDANNEIMSQPNSTGYTIDNSLNOIVNANRANISVC
GNDDATARYKDIKLYASDALSNOPTIRNLQTLANAPRTTISIANONNNFNQGDLS
INOSISDLINQIVNFCRPEVDVYNGGEVWCQFEPKPIQGLMAAAGLGLFLFQ
STLLVTRKMTNSDILAMNYRDKNOSTSKYSTKRIKQDVRTEELMARPIL
SSTSNTEIRDLERVRPDMDIVOPSEYVSYGTGSRKHYPRQVLPDRHQMSPNY
DIOQYGTATRFNDILDFHOSNGSTAGRL"
join(13937..14329,14377..14533,14662..14822,14873..15040,
15203..15286,15367..15528,15577..15697,15742..15953,
18567..18749,18794..18874,18928..19155,19478..19573)
/gene="M28.2"
/protein_id="CA90130.1"
/db_xref="GI:3878694"
/translation="MKLCFRGISFTSIARCANSKPKPOACCKSSLETFRPMPIS
GSTGYGKGISDGLQKAPDSOTYODLNNISKREKSTADITKSNVSQVAILFWSY
GTFILATIACTVTCIWHFCACAPKSAVSRTGVTGGLATLVASIVAILF
NLAETLDFDSIDNGVTADKIVTDIFDLVSGNKLQCEVSSQALAEIRSIKYN
AEDVIDGTTRVGLTAVNNPDSDAFKRSTEVYKNGVYLIDGLQTLQMSPTGPDCHN

```

```

TRANLEKVIITIOSALSALSLAHSFNSSELEINDLIKIKIDEVESQKGAAGIAD
SAEVSDDDTINSIVFEDKIQDEKAVAKTISIPSIHQKVSYSTGAVTIGRVITL
ACGLICCIILAIWVYVLSMKQDGLAKKLSIVAIISFVSIIVLICILFKIAMVFAAG
LVSVSVCPVPGGDKYTLVYNSDSSNMKSFNOHNTAKKLLDNDEQAASNLATCD
NNNDOTFEIVSESLSMSHLSOOFPVKLETLLENAPKTSIKDLNKTYETTEAL
KOAVTKFLKLODDIEKCRPLDVYNNAGISVCEQGLPHGMASVGLCGVLEFQI
ILLITFRMLRTSKNTKVNKNGSKSDRTGTSQSKKSKKKKKKRNKAPTDK"
join(20244..20468,20516..20698,20746..20902,21009..21169,
21220..21378,21607..21690,21739..21897,21948..22068,
22115..22337,22457..22643,22712..22894,22941..23021,
23067..23285,23437..23531,23579..23791,23862..23997)
/gene="M28.9"
join(20244..20468,20516..20698,20746..20902,21009..21169,
21220..21378,21607..21690,21739..21897,21948..22068,
22115..22337,22457..22643,22712..22894,22941..23021,
23067..23285,23437..23531,23579..23791,23862..23997)
/gene="M28.9"
/translation="MKLRLEVFIALISIAFCAKSPKPOASGPAPACPKSSEVEKPM
KPELTSVGFYGTGKGISDLSQKTPDSQEDLNLSGKSEYGETLSMILSOVTAI
VMSYGIYVLIITIAALCIYTCIMQCCSTCAPRESAVRSEYTGMYLALLATLATTLA

```

Query Match 4.6%; Score 53.6; DB 3; Length 40265;
Best Local Similarity 51.4%; Pred. No. 0.048;
Matches 130; Conservative 0; Mismatches 139; Indels 3; Gaps 1;

QY	480	TGAGCAACCGGTCACACAGTTTCATCTATCGTGGCAGCAGTTATCATTAAG	539
DB	5385	TGATCAGAAATGGAACATCTGTTTCAAAACGCTCTTGTGGAGACGTGGTCACATCAAG	5444
QY	540	GACATCCGATTTGAAACCGTGTGATCTTGCAGCGGTGTCATTCCTCGTTGCGA	599
DB	5445	GGCTCTTAAAGAGGTTTAACTAAATAAATAATGTTGGGCAATCATGTATGTGA	5504
QY	600	TGATGCTAACGGTGATACGTGGAATTTGAATGCTGATGATGCTCTTGATTAATA	659
DB	5505	AGCGGAGCAGGTCACAAATTTGAAGTATGATGATCAACATGATGACCTTGATCAAT	5564
QY	660	TTTGTAAATTAATTTGGAATATCCAAACAGATTTAATGGCTGCCAA---GAAGCTACAGT	716
DB	5565	GATGCTTCAACACCAACATATTTGAGATGAGATGATGATGACACAGTTGATGATAT	5624
QY	717	ATACCAATATGCGGATGATGATCAACGCTTTTATCAATGCCAGATCAGTATT	768
DB	5625	ATTCAAGTTTCTGTATGATGATCAACCGTTGATTTGATGATGATCAATTAATTTT	5676

RESULT 30
CEM06D12/c
LOCUS CEM06D12 34368 bp DNA linear INV 24-JAN-2002
DEFINITION Caenorhabditis elegans cosmid W06D12, complete sequence.
ACCESSION Z82073
VERSION Z82073.1 GI:2653149
KEYWORDS HMG.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditoidae; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1
AUTHORS none.

TITLE	Genome sequence of the nematode <i>C. elegans</i> : a platform for investigating biology. The <i>C. elegans</i> Sequencing Consortium
JOURNAL	Science 282 (5396), 2012-2018 (1998)
MEDLINE	99069613
PUBMED	9851916
REMARK	The <i>C. elegans</i> Sequencing Consortium.
REFERENCE	2 (bases 1 to 34368)
AUTHORS	Basham Y.M.
TITLE	Direct Submission
JOURNAL	Submitted (06-NOV-1996) Nematode Sequencing Project, Sanger Institute, Hinxton, Cambridge CB10 1SA, England and Department of Genetics, Washington University, St. Louis, MO 63110, USA. E-mail: jlesaner.ac.uk or rw@nematode.wustl.edu
COMMENT	On Nov 29, 1997 this sequence version replaced gi:1695114. Coding sequences below are predicted from computer analysis, using predictions from GeneFinder (P. Green, U. Washington), and other available information. Current sequence finishing criteria for the <i>C. elegans</i> genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note. IMPORTANT: This sequence is not the entire insert of clone W06D12. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions. The true left end of clone W06D12 is at 553 in sequence A1023857. The true right end of clone W06D12 is at 2303 in sequence Z93782. The start of this sequence (1..104) overlaps with the end of sequence A1023857. The end of this sequence (34265..34368) overlaps with the start of sequence Z93782. For a graphical representation of this sequence and its analysis see: http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=W06D12 IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions. Location/Qualifiers
FEATURES	1..34368
source	/organism="Caenorhabditis elegans"
gene	/db_xref="taxon:6239"
CDS	/chromosome="V"
gene	/clone="W06D12"
CDS	join(1055..1180,1224..1326,1402..1631,1673..1816)
gene	join(1055..1180,1224..1326,1402..1631,1673..1816)
CDS	join(1055..1180,1224..1326,1402..1631,1673..1816)
gene	/gene="W06D12.6"
CDS	/codon_start=1
gene	/protein_id="CAB04927.1"
CDS	/db_xref="GI:3880551"
gene	/db_xref="SPRMBL:045895"
CDS	/translation="MEOLETCDLTKFTPSPENPTEMRYVETIPGVSLADYLSIGPQFTMAGYVCAKGMHNGFALFVPEKRSITTRHCNCPVSGVFTYQHSPTNQHTRAIDREDVITNRTIEGTRKNIVSECFYTCAGDGLVVFSENGPTIPVRFPGQITAKCRPPVSHSDVMKPFYENWMMFDNMYMORPOITCFEK"
gene	join(2263..2429,3400..3736,4267..4764)
CDS	join(2263..2429,3400..3736,4267..4764)
gene	join(2263..2429,3400..3736,4267..4764)
CDS	join(2263..2429,3400..3736,4267..4764)
gene	/gene="W06D12.7"
CDS	/codon_start=1
gene	/protein_id="CAB63322.1"
CDS	/db_xref="GI:6580250"
gene	/db_xref="SPRMBL:090346"
CDS	/translation="MTSISNNENPFTYIMTROLIIFGLTFEYPLFIYIVLCRTPKQCGNLKMSCLFPHFAATCOMLCSFLIIDFTYTPSPKITYTIDGILIDPIDAKYIYTSMTIISALTSKSLFCHSVNIFEMLMKNSWMTSEEMLMTVLMFPFMSLIGFTIRQDPNVAKLTRQJSSYTPDCIDNPNTVPLIPDAEGKSLVFFLMFLFMATASLISADILSHLSRMTNOSDKTRKMKHOKFNRSILFOVLIDISTSPFYIISNFATLQMFOMPELT YFVDFMSKSGSTACIFVIFWYDYPYRKWLQELIYARSKSTQTSVRHVIYIAVMDRGNEP"
gene	complement(join(5022..5258,5402..5535,5584..5957,6431..6584,6629..6770,6823..6945,7344..7618,7718..7908,8766..8883,9010..9145,9196..9207))
CDS	/gene="W06D12.5"
gene	complement(join(5022..5258,5402..5535,5584..5957,6431..6584,6629..6770,6823..6945,7344..7618,7718..7908,8766..8883,9010..9145,9196..9207))
CDS	/gene="W06D12.5"
gene	/note="cDNA EST yk224a2.3 comes from this gene"
CDS	cDNA EST yk224a2.5 comes from this gene"
gene	/codon_start=1
CDS	/protein_id="CAB04926.1"
gene	/db_xref="GI:3880550"
CDS	/db_xref="SPRMBL:045894"
gene	/translation="MTSEQNNYTPKGNVPRHSAVLEIHEEEDSTIPEEDTLENKPSHSENDFKRNAIRSEETPISEIORIEDPLASDSEENKEKPARISVSNADLEERHRDLSIPENDRSQSRYSRHSIDKGLKONNSEVDYEVYERNNYTVGRPRPFYNKLPGRQHRSVHTDASRFSFGAMGEVPTLRFRDKNWFAFHROIGFRHESVVILVILYTLGLGAMFWTVAESRREHAKTLDHVNLEHLDRLANTESEVNNITNTTTEEMKYIRARATIELMKLBQYKSTYYKLEADNNKWTRESAFPSSMNYTTTGGSIAPESTIGOVLCVGVGTFVPTLVLRDGLGFVHLTKLVHGIQKRELGKHHVDEDEIISLPKACLLASLYGACTFEYFDLSGPEPTGMDMLCEYFSEISLSTIGLGDIMPNNATVGNRRNITYENKIFAPISITIFFGMAVTVVRNMEYIAVENGIFGATLVENKLDALVTRSSASVREDEPSPTRKVGORALSDVAGSEODETPHEINLNTLRSIATFMKSMADLYGGGFGVQLRBDLHSDSQNMQTVMSQQLHRRATASQSQPQA TGCNNNV"
gene	join(13888..14185,14708..14777,14968..15102,15147..15423,15938..16234)
CDS	join(13888..14185,14708..14777,14968..15102,15147..15423,15938..16234)
gene	/gene="W06D12.4"
CDS	join(13888..14185,14708..14777,14968..15102,15147..15423,15938..16234)
gene	/note="contains similarity to Pfam domain: PF01461 (7TM chemoreceptor), Score=-36.9, E-value=5.1e-13, N=1"
CDS	/codon_start=1
gene	/protein_id="CAB04925.1"
CDS	/db_xref="GI:3880549"
gene	/db_xref="SPRMBL:045893"
CDS	/translation="MTVSFEVSDNLTIAVITCVSCNILLIYLVTKSKRHGTTRYLMFTFALNGILHSVFQDTKPIFLMDGNKHSILVYFLNRPFSKMAVLAISEYGFELAVIFAVHFIETRWYITIKNDKIRFENPFCWIGISYGVGEYSYLHVEFSERPDKTEEFIRKMGSEDFGHTMDIEYMAARFYKSEEGALIDRQGVGFYAAKRTMLSPILLIYCGAIFAIKTRTKRDKLSRKSRELEINLELALIVQVIVIPLMYICPFLMMNFPFLIGETARLTGMGVALYPGIDPLAIFIIYVANNRRALKRFRKVELVQPSSTVRRDISTSRALQVAGNOYAPAVVIN"
gene	join(19946..20159,20215..20551,21208..21658)
CDS	join(19946..20159,20215..20551,21208..21658)
gene	/gene="W06D12.3"
CDS	join(19946..20159,20215..20551,21208..21658)
gene	/note="contains similarity to Pfam domain: PF01069 (paty acid desaturase), Score=465.8, E-value=1.1e-136, N=1"
CDS	cDNA EST yk148b12.5 comes from this gene
gene	cDNA EST yk412h2.3 comes from this gene
CDS	cDNA EST yk412h2.5 comes from this gene
gene	cDNA EST yk548c4.3 comes from this gene
CDS	cDNA EST yk197g12.5 comes from this gene
gene	cDNA EST yk548c4.5 comes from this gene
CDS	cDNA EST yk661f6.5 comes from this gene"
gene	/codon_start=1
CDS	/protein_id="CAB04924.1"
gene	/db_xref="GI:3880548"
CDS	/db_xref="SPRMBL:045892"
gene	/translation="WFOIKVDAIITSKFOFLADINETIROMESEKQVYIKETIYKQVVALFVALHIGALVGLYQVFOAKMAVTVGVFLHTLIGSGVYGGAHRLMAHRAVATLSMRFVIMLINSIAFQNDIIDVARHGRCHKMTDADPSTHGRGFEHAKMLVYKRLDLMKIOGGLIADSDLEDYEDPVLMEFORKNRLPLVIGICFAPLPIVPLVMEGSAFLAFYAALEFRYCFILATWCIINSHWGQOPYDHOSSNDWTSLAAEGEGSHHFMFPDQYRTSEHAEFLNMTKRVLLIDFGASIGMVYDKRTPEVIOCKKSGCEPRKMKHKG" complement(join(21764..21883,22445..22683,22761..23145,24058..24140,24197..24325,24883..25051,25103..25312))
gene	/gene="W06D12.2"
CDS	complement(join(21764..21883,22445..22683,22761..23145,24058..24140,24197..24325,24883..25051,25103..25312))

DEFINITION Caenorhabditis elegans cosmid K06A1, complete sequence.
 U23449
 VERSION U23449.2 GI:14574206
 KEYWORDS HTG.
 SOURCE Caenorhabditis elegans.
 ORGANISM Caenorhabditis elegans.
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
 1 (bases 1 to 26449)
 Waterston,R.
 TITLE Genome sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium
 JOURNAL Science 282 (5396), 2012-2018 (1998)
 MEDLINE 99069613
 PUBMED 9851916
 REFERENCE 2 (bases 1 to 26449)
 AUTHORS Fulton,L.
 TITLE The sequence of C. elegans cosmid K06A1
 JOURNAL Unpublished (2001)
 REFERENCE 3 (bases 1 to 26449)
 AUTHORS Waterston,R.
 TITLE Direct Submission
 JOURNAL Submitted (13-JUL-1995)
 REFERENCE 4 (bases 1 to 26449)
 AUTHORS Waterston,R.
 TITLE Direct Submission
 JOURNAL Submitted (25-JUN-1998) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 REFERENCE 5 (bases 1 to 26449)
 AUTHORS Waterston,R.
 TITLE Direct Submission
 JOURNAL Submitted (28-JUN-2001) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA
 6 (bases 1 to 26449)
 Waterston,R.
 TITLE Direct Submission
 JOURNAL Submitted (23-MAY-2002) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA
 On Jun 28, 2001 this sequence version replaced gi:733545.
 Submitted by:
 Genome Sequencing Center
 Department of Genetics, Washington University
 St. Louis, MO 63110, USA, and
 Sanger Centre, Hinxton Hall
 Cambridge CB10 1RO, England
 email: rwenemate@wustl.edu and jessesanger@ac.uk

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m13 subclone.

For a graphical representation of this cosmid sequence and its analysis see:
<http://www.wormbase.org/db/seq/sequence?name=K06A1;class=Sequence>

NEIGHBORING COSMID INFORMATION

The 5' cosmid is K03H9, 900 bp overlap; the 3' cosmid is T28D9, 2500 bp overlap. Actual start of this cosmid is at base position 897 of K06A1; actual end is at 16582 of T28D9.

NOTES:

FEATURES

source

Coding sequences below are the result of integration and manual review of the following data : computer analysis using the program Genefinder (P. Green and L. Hillier, personal communication), the large scale EST projects of Yuiji Kohara (http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html) and The C. elegans ORFeome cloning project (<http://worldb.dfci.harvard.edu/>), similarity to other proteins from Blastx analyses (<http://blast.wustl.edu/>), sequence conservation with C. briggsae using Jim Kent's MABA alignment program (Genome Research 10:1115-1125, 2000), individual C. elegans GenBank submissions, and personal communications with C. elegans researchers. tRNAs are predicted using the program tRNAscan-SE (Lowie, T.M. and Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).
 Location/Qualifiers
 1..26449
 /organism="Caenorhabditis elegans"
 /strain="Bristol N2"
 /db_xref="taxon:6239"
 /chromosome="II"
 /clone="K06A1"
 98..3998
 /gene="K06A1.3"
 /note="for a graphical representation of this gene see: <http://www.wormbase.org/db/seq/sequence?name=K06A1.3;class=Sequence>"
 join(98..182,258..358,462..536,656..798,896..1277,1325..1780,2763..3198,3244..3494,3687..3773,3819..3998)
 /gene="K06A1.3"
 /note="similar to C. elegans cuticulin precursor; coded for by the following C. elegans cDNAs: YK334a2.3, cm20c2, YK60c1.5, YK334a2.5"
 /codon_start=1
 /product="Hypothetical protein K06A1.3"
 /protein_id="AAC24298.2"
 /db_xref="GI:14574207"
 /translation="MTSTDFLLMLALFNSAQPIVEPDNHSIDPEVSGHSGFMSLKV NTEKSPSHVFKGHFRKDGSEFNTAATFEFSKCDVMRQREANPKMATATVAVQ LHPITPTVDAVYLKRCFKAEKAVGAEVSDPTPVQLDESGPYCSYTIKESF NGPIAKFAGDGLVLYHWECESEAVOMVYNDVPGYEGEKYGEKSGCEDYIMP NLIVENRTKAFVNSNAPFPDQNNVRISCTISCPITSGTCHPKCDASASTDDAS EPIGHELRDELITPMTSSSEITPSTTSTTTTSTTTATTEPTTTEPTTTEPTTVO VADPTITVLTPTDPTDSTIRFLSKPSKSNNTVWEDKESGIEISBGAHAQAQHT SSEBPTTKREIKRKDATVDVDEHQIISYLPAVSADVILNARCFVWISLKLFL LKVCYVYFSSRHAGRPRTQSMOPFHYRGLYRNSKRSLSGSSNOCINNQNNGO CTSDSCFHFATSEGDIERFDGSOKEPKLSTLSCAVLINTDSSVCLGARPTHC CTSSSPKICAPKERRKMAELIOLMTMSQVMVLVGVHGLSCVLLVTFYMKSVKYL P KNYIYIPESBMTITLAKIKFDAGIGSLYSFPLIPIPFPSLFIYKKDKYINKTC ALFLAIFKLTIVPIFYECFLSTVCLMHRDHPGICRLNGAIWQFMVFR"

gene

CDS

join(4064..4237,4286..4362,4415..4490,4535..4599,4644..4803)
 /gene="K06A1.2"
 /codon_start=1
 /product="Hypothetical protein K06A1.2"
 /protein_id="AAC24297.1"
 /translation="MLLVITQYLLSKLKPNDIPRISEETPSQVAPDTILSTOSGEMR IESFENEMTVTHPRPFVPRVLGILILWKNRNDMLALRALSTGCFYKLYPSKEFLI PGKEISIEVEQCTRAESPITLNHITLIEWPLDSTPTPGDGLRMSKPYLVPSHMWY FVLVYVDDSVTSVSPSHVSSE"

gene

CDS

complement(join(5030..5039,5091..5185,5232..5383,5430..5545,5594..5725,5773..5969,6018..6415,6652..6983,7072..7234,8113..8457))
 /gene="thr-22"
 /note="for a graphical representation of this gene see: <http://www.wormbase.org/db/seq/sequence?name=K06A1.4;class=Sequence>"
 complement(join(5030..5039,5091..5185,5232..5383,5430..5545,5594..5725,5773..5969,6018..6415,6652..6983,7072..7234,8113..8457))
 /gene="thr-22"


```

/note="similar to C4-type zinc finger domain (exon 2);
coded for by the following C. elegans cDNAs: yk325g8.5,
yk324h4.5, yk324h4.3, yk325g8.3"
/codon_start=1
/product="C. elegans NHR-22 protein (corresponding
sequence K06a1.4)"
/protein_id="AAC24301.1"
/db_xref="GI:3253111"
/translation="MSTDLIEYFVFPPTDLFRNFPEDVPPSPSLDVGDSLGSSPSVD
LGFQPSFDIDFCIKWKGSSSPPESSNHLPEITVKAISTSPGMFSPILDSRCHVCSF
TANTLHFHGSGSCACAAFFRRSVSMSTECIGTGDTPCRRHYELRMICRCRRTIK
CIDAGMRRELYQARKKEETRYARRSGLVYKSKEDDVESSEYVNTYINVDSSPK
MESGIMENTIPESYLSPPSSSOPLDMTVTPPLSTPESILTPATGDMHCORVLS
NSPPDIHOASTSHAGFIOHVQFVPEVVENKTFFELVDHYRTEASLNDPRKIMYDT
OTRDVPTTCECEYENHOLKRFVYKFCGVRKIDYPMIIDDYVNOFPERQALHNDAKV
VTRMACAVDSMLASAYSTYKVGTEKRLIFNCDYINMNPISGDEPGAGTEFQPO
EHEKRTMLPKLKOYFDLAIIPARLEVSFEELVLLKALIMQISNRULEERRALCA
ROBDITVOALHKVVERGDEDPAIRVGOLLISMSYITEQVQAMTNSYLVMTFFDVVSC
DSIMYDLISFRDD"
/complement(9832..11847)
/gene="K06a1.5"
/note="for a graphical representation of this gene see:
http://www.wormbase.org/db/seq/sequence?name=K06a1.5;class
=Sequence"
/complement(join(9832..9922,10394..10716,10769..10943,
10993..11636,11686..11847))
/gene="K06a1.5"
/note="Contains similarity to Pfam domain: PF00400
(G-beta), Score=70.8, E-value=9.6e-18, N=5"
/codon_start=1
/product="Hypothetical protein K06a1.5"
/protein_id="AAC24299.1"
/db_xref="GI:733550"
/translation="MANVYRMKSKNDQDIDANKRLADESRYSLVSSQDKRLRIETD
AIVKMTVLENEIAELKEENSVINTERVSLATCNLTKEKTLQDNERFQLNRIEL
OEKSAEFMNAEILQEBRQOMRIEIOAKATADLNGDERASSAFGSPPTLDEPMAT
DYLSPSEVKRLSTHGDGEVHVMWSDPTFATAGSDSKVOIWRYSPPKTAISKYTSIG
CICPVNRLDYDSORHYCLASSNDKTRKLNIDSORLLSTFSGITDKVSSRLQSHNV
ISSADNRITKMDIISITKCLKSTLVGTVFDIVAKGVSSOSITSHFDKVFVWAR
SSDATYSELGQVSSLDISMDGLQVLASRDYDLSLIDVRNGLIHLVSAEQYKTCF
DSTRATFSSGVEYVLAGSSNSVFVINTYTKTLKRYVTKARSDSAOIMSLANMPSGRG
LLACDRQKTCCTLMR"
/13643..15109
/gene="K06a1.1"
/note="for a graphical representation of this gene see:
http://www.wormbase.org/db/seq/sequence?name=K06a1.1;class
=Sequence"
/join(13643..14066,14112..14328,14377..14495,14671..14939,
14993..15109)
/gene="K06a1.1"
/note="Similar to transcription factor AP-2
(SP:TAP2_HUMAN, P05549)"
/codon_start=1

```

Query Match 4.2%; Score 49; DB 3; Length 26449;
 Best Local Similarity 59.9%; Pred. No. 0.59; Mismatches 55; Indels 0; Gaps 0;

```

Matches 82; Conservative 0; Mismatches 55; Indels 0; Gaps 0;
QY 271 AATCCAGGTGATATTTTGTACAAACAAGTGTGTCATTTGTTCAATTCATTAATTTGTT 330
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 659 AATCCAGGAAGAAATGGCATATACAGCTAGTACTAGTTCAGCTTCACTCAGCTTTTACA 718
QY 331 ACCAAAGTTATCGTCAATTCAGAGTACATGCTTTTACATGGAAGCTGATAAAAGCTT 390
      || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 719 ACTAAAGTCAGAGAGCTATAACTTCGTGTTTCTACAAAGAACCGGAAAGCAGATC 778
QY 391 AGTGCACAGATTGAGT 407
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 779 GGTGCTGAGCTTTCTGT 795

```

RESULT 33
 U21318

LOCUS U21318 31731 bp DNA linear INV 23-MAY-2002
 DEFINITION Caenorhabditis elegans cosmid K03H9, complete sequence.
 ACCSSION U21318
 VERSION U21318.1 GI:687828
 KEYWORDS HTG.
 SOURCE Caenorhabditis elegans.
 ORGANISM Caenorhabditis elegans.
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
 Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
 1 (bases 1 to 31731)
 Waterston,R.
 Genome sequence of the nematode C. elegans: a platform for
 investigating biology. The C. elegans Sequencing Consortium
 Science 282 (5396), 2012-2018 (1996)
 99069613
 MEDLINE 9851916
 PUBMED 2 (bases 1 to 31731)
 ANDERSON,K.
 The sequence of C. elegans cosmid K03H9
 Unpublished (2001)
 3 (bases 1 to 31731)
 Waterston,R.
 Direct Submission
 Submitted (13-JUL-1995)
 4 (bases 1 to 31731)
 Waterston,R.
 Direct Submission
 Submitted (23-MAY-2002) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 Submitted by:
 Genome Sequencing Center
 Department of Genetics, Washington University
 St. Louis, MO 63110, USA, and
 Sanger Centre, Hinxton Hall
 Cambridge CB10 1RQ, England
 email: rwenemate@wustl.edu and jess@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.
 It may be shorter because we only sequence overlapping sections
 once, or longer because we provide a small overlap between
 neighboring submissions.

This sequence was finished as follows unless otherwise noted: all
 regions were double stranded, sequenced with an alternate chemistry
 or covered by high quality data (i.e., phred quality >= 30); an
 attempt was made to resolve all sequencing problems, such as
 compressions and repeats; all regions were covered by sequence from
 more than one m13 subclone.

For a graphical representation of this cosmid sequence and its
 analysis see:
<http://www.wormbase.org/db/seq/sequence?name=K03H9;class=Sequence>

NEIGHBORING COSMID INFORMATION

The 5' cosmid is T25D10, 200 bp overlap; the 3' cosmid is K06a1,
 900 bp overlap. Actual start of this cosmid is at base position 1
 of K03H9; actual end is at 8345 of K06a1.

NOTES:

Coding sequences below are the result of integration and manual
 review of the following data: computer analysis using the program
 Genefinder (P. Green and L. Hillier, personal communication), the
 large scale EST projects of Yui Kohara
 (http://www.ddb.jng.ac.jp/c-elegans/html/CE_INDEX.html) and The C.
 elegans ORFome cloning project (<http://world.dcc.harvard.edu/>),
 similarity to other proteins from Blastx analyses
 (<http://blast.wustl.edu/>), sequence conservation with C. briggsae
 using Jim Kent's WABA alignment program (Genome Research
 10:1115-1125, 2000), individual C. elegans GenBank submissions,
 and personal communications with C. elegans researchers. tRNAs

similar to SWISS-PROT:P34029 (DP3B_SPIC1) Blastp2 p-3e-14
C-43%
/codon_start=1
/evidence=not_experimental
/transl_table=4
/product="DNA POLYMERASE III, BETA CHAIN"
/protein_id="CAC13175.1"
/db_xref="GI:14089415"
/translation="MKLFLEKSRFLRPLENTTKITIESNNSNYELRCAYFVYRQDKIL
ISSNEDISINNEYLLDGIKNTGSESDYVLNSLKNIKKSDSIILEKKTPLKE
SDQSYEINILNLDREPDIKFGLNNSQLKTFQFQALKNVFPADLSNAOELLIS
VNLSTKNNILTLVATNRAIAMQRIKEDSOEENPLINSKVKELISLSMNTLISP
GTEELKISGNLEIKTEVIEIPYMANVENFVPEKNDYFVHIDKKEILSLDKSVLND
KNGKILIEVNPSEKKEKKLSSYWDLGFSEVNDPEVESEELKLFITNAVYAKES
INFDGKISIFITENKDRVYISSETNLNKKLIAALRGH"
2864..3085
/gene="MYPU_0030"
2864..3085
/gene="MYPU_0030"
/note="identified by sequence similarity; putative:
similar to SWISS-PROT:P05650 (YAA_BAC50) Blastp2 p-6e-06
C-45%
/codon_start=1
/evidence=not_experimental
/transl_table=4
/product="conserved hypothetical protein"
/protein_id="CAC13176.1"
/db_xref="GI:14089416"
/translation="MTLKGFLITGOLIKRIKLIDSGQIKQFMOSHOKVINGKLI
NTRNTYRVGDTIWDQLYTIKRAQNNEE"
3088..3771
/gene="MYPU_0040"
3088..3771
/gene="MYPU_0040"
/note="identified by sequence similarity; possible:
similar to SWISS-PROT:P43055 (YL1_MYCH0) Blastp2 p-1e-28
C-53%
/codon_start=1
/evidence=not_experimental
/transl_table=4
/product="conserved hypothetical protein"
/protein_id="CAC13177.1"
/db_xref="GI:14089417"
/translation="MKYIKSIHKYVPRKLSHYKLFYGFHNRTYLGFGYVQIRL
KINDIYNHDEYKYLKNNEDYLFYKNKRVYTKKPFDEHFRANKESLKNLDIEKES
QNLDLAKMNLFYTKRDPKYLSLVERKDFNDVHGLNFKNMLANDKNEVCLIDE
WIRKGSYFDIICIKYKNFNKNELKRPDLNEKFPDYLMSYKFEIEAEKRYYSKM
SDSFWEYERK"
complement(3778..4422)
/gene="MYPU_0050"
complement(3778..4422)
/gene="MYPU_0050"
/note="no similarity found"
/codon_start=1
/evidence=not_experimental
/transl_table=4
/product="unknown; predicted coding region"
/protein_id="CAC13178.1"
/db_xref="GI:14089418"
/translation="MERNVSEFVQICEKNLFLYSLAFETLSMNGEKIDOKKHIPQ
VFMSIDSYRLKKKFNVNINSNINFLKPRVYLDKSNMINEVSEFVILIPST
LKKIKSPKSLKNILOYINKNKEIKRSSLFKKIPNFPINIKAIYSISKHOGF
VLSNVNENTFEKNMFLNITFTKELIQFLNLTFRVIEELSHFKRNYKDGYKIIIF"
4603..6759
/gene="MYPU_0060"
4603..6759
/gene="MYPU_0060"
/note="no similarity found"
/codon_start=1
/evidence=not_experimental
/transl_table=4
/product="unknown; predicted coding region"
/protein_id="CAC13179.1"

/db_xref="GI:14089419"
/translation="MKKSPKRNFKSSKSVYDKKTKRNNYILGSGIFTLIVGVVPIE
AVNSTAQRPALQEDAPLNFINGRNSDQYTSYDVSILNGVGLKSEYFKQDEVFTRKSN
AYSEFLYDEGRKASENLANFVKEGSSFPSEFTNLKIDELIRASERELDAKRN
IQLOOGPANNEAEWLKQIRENDKYGKKNEDAEFLYQKRIKNDAYRDPQILSL
NVDLKKVASRDLIDHDNTKVFIRKQGPLPEFDFDNTKQNVFVPSNDSNTSAV
YTPRSFIFNEHRNENLAKRYLERGKTIYSSTLLTPKSDAKRESEYVSKDFTL
ISYSAFNPQDLQDATTNNNGVDLLKAKNAYGVETLSENISAFETROADIDIOHIT
VAYNSKNGRGFGVGSGLDVSITKIDSLTLTGITFVAELDATNINPLDTFIE
KVENFEGLLPDAPPNVSDAKTNRELSOKIOWSKREPFIRIKRIEATFTPDRN
GREILQKRAMAYKKADTDNNEVYVLDGLNIVKHEVSTMDENNYLNDLYSL
HSSSTORVGEDPDKILKYSDNDFPAQGFESNAEFKPYREKYSNLTVMFNSERN
ARAFPLNTNHNDSNOSKEVERKRFIESEKQNTYVELDENIEFLKAKPNDINK
OSPREKVDKMSFFENILIPMPFDVGSRLKQDN"
6759..8189
/gene="MYPU_0070"
6759..8189
/gene="MYPU_0070"
/note="identified by sequence similarity; putative:
similar to TREMBL:Q49557 (Q49557) Blastp2 p-3e-15 C-29%
pIscon: pos. 18-28 PS00013 | PROKAR_LIPOPROTEIN
prokaryotic membrane lipoprotein lipid attachment site"
/codon_start=1
/evidence=not_experimental
/transl_table=4
/product="P60-LIKE (Mycoplasma hominis) LIPOPROTEIN"
/protein_id="CAC13180.1"
/db_xref="GI:14089420"
/translation="MKIKRIKILKTLPFTSTPIITSSILISCGOVIDSTERKINDET
KITSKESVNLVYEVIVSOYSNDSIDFTKNDOSLDKNSOYKDAFLAKRYVSN
ERLKSTSVNYDORANMEKANTSVNKTOSITJLNNNGHPLDENFVFLNKTSPYR
HEITKMLMVKVNYLINTSEBOVKKSPFPEYKRIKILDKNNDKRTYESIDSKNPFEL
KYLKDKRMITMKMYKSETDSKVLAIDVYINIQFENENVOYOSVESLDTEDISL
RLNDEYKRALETONSRLNPLIRAFETVETTEAGGLANNIDOLKQKEVINGEFLAT
KNKTIISKNOLLNOSQSGDGLREFLOITNFKGKSEIMSYLTKLPPEEDYATD
AEYKDKDQSDPTIKRGFTFEKFNKPMQDAESANETLKLAYWLYPDNLSLFOAKNH
FSSIGYKIESNNPIITRLKNEGTLK"
8190..8540
/gene="MYPU_0080"
8190..8540
/gene="MYPU_0080"
/standard_name="hit"
/note="identified by sequence similarity; possible:
similar to SWISS-PROT:P32083 (YHIT_MYCH1) Blastp2 p-7e-22
C-56%
/codon_start=1
/evidence=not_experimental
/transl_table=4
/product="HIT-LIKE PROPEIN (CELL CYCLE REGULATION)"
/protein_id="CAC13181.1"
/db_xref="GI:14089421"
/translation="MSDNHSIFPKIIRDRPAQIYEDDOAIAIMDKPYPNGHFLV
IPKHSNLDKIDIEOSINHLMKIAIKLAKKIEKKEFEDEKFLIINNEKAGOVVYHTH
IHIFPRKKEGHEH"
8643..9296
/gene="MYPU_0090"
8643..9296
/gene="MYPU_0090"
/standard_name="p1k"
/note="identified by sequence similarity; possible:
similar to SWISS-PROT:P75550 (RUL1_MYCPN) Blastp2 p-6e-36
C-44%
/codon_start=1
/evidence=not_experimental
/transl_table=4
/product="50S RIBOSOMAL PROTEIN L11"
/protein_id="CAC13182.1"
/db_xref="GI:14089422"
/translation="MEVDQWLTLPHLEIRIVLIMAKEVYKIAQLQFIAGAKPGPS

Query Match 4.2%; Score 48.4; DB 1; Length 327650;
Best Local Similarity 46.7%; Pred. No. 0.77;
Matches 154; Conservative 0; Mismatches 176; Indels 0; Gaps 0;

OY	108	TTCAATTAACAATCAATTTTAAATACACGTAATGCATTTCCAAAGACATGTTATGTCAAAAG	167
Db	52535	TACAAAAAATAATAGATAGATTTACTACTATTTGGATACACCTCAATATATCTCGCTTTTGT	52594
OY	168	TCCTTATATGATCAAGAAGGTTGCCGTAATGATGAAGGTGGACGCTCAAGTTGCCGGAATTTTC	227
Db	52595	TGGCTAGATTTTACCTGATTTATAGTAATATGAATTTATGACAAGCAAGGAGGAGGATTTTC	52654
OY	228	ACTTCATTTTGGTTCATGCAATGTTGGCGGTACAGATCTCGAATCCAGCGGATATTTC	287
Db	52655	AATAGCATTTGTTTATATATATATTATGGACATTTTCTATTAACTACATATGGATTTTTC	52714
OY	288	TGTAACAACAACCTGTTTGTCAATTTGCTTTCAATCAATATTATTGTTACCAAGTTGATCGTGC	347
Db	52715	TGTTCCAAATTTTGTGTTTATATGCTTTTGGGCAAAAAAATACTTTTCAAAATTAATCTACG	52774
OY	348	ATATGAGATACAAATCTTTTACATGGAAGGCTATTAACAAGCATGTAAGTCACATATGAGGT	407
Db	52775	AATGTTTTTTTGGAGATTTTTCATATTCGACATAGTATTTTTTGTTTTAAATTTACATCTCTTTC	52834
OY	408	ATCTGAATAATCACAACTGCTTTTCCAAGTCA	437
Db	52835	ATCTAATAGTATCCTTCATGCAACAAGTAA	52864

RESULT 35				
AC116984				
LOCUS	AC116984	53150 bp	DNA	linear
DEFINITION	Dictyostelium discoideum chromosome 2 map 30006030-3053778 strain AX4, *** SEQUENCING IN PROGRESS ***; in ordered pieces.			

VERSION	AC116984.1	GI:20042947
KEYWORDS	HTG; HTGS; PHASE2.	
SOURCE	Dictyostelium discoideum.	
ORGANISM	Dictyostelium discoideum	
REFERENCE	Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.	
AUTHORS	1 (bases 1 to 53150)	
	Gloeckner, G., Eichinger, L., Szafranski, K., Pacheco, J., Dear, P.,	

TITLE	Sequence and Analysis of Chromosome 2 of Dictyostelium
JOURNAL	Unpublished
REMARK	The Dictyostelium Genome Sequencing Consortium
REFERENCE	2 (bases 1 to 53150)
AUTHORS	Baumgartl,C.
TITLE	Direct Submission
JOURNAL	Submitted (05-APR-2002) Genome Analysis, Institute of Molecular
REFERENCE	Biotechnology, Beutenbergstr. 11, Jena 0745, Germany
	3 (bases 1 to 53150)

JOURNAL Submitted (18-MAY-2002) Genome Analysis, Institute of Molecular Biotechnology, Beutenbergstr. 11, Jena 07745, Germany

COMMENT CDS predictions from Genemid may contain errors. Further information is available from IMB Jena, Department of Genome Analysis (<http://genome.imb-jena.de/dictyostelium/>) and the University Cologne, Institute for Biochemistry I (<http://www.uni-koeln.de/dictyostelium/project.shtml>)

Funding Agency : Deutsche Forschungsgemeinschaft (DFG).

* NOTE : This is a 'working draft' sequence.

* This sequence will be replaced

* by the finished sequence as soon as it is available and

* the accession number will be preserved.

```

FEATURES
source
location/Qualifiers
1..53150
/organism="Dictyostelium discoideum"
/screen="AX4"
/db_xref="taxon:44689"
/chromosome="2"
/map="3000630-3053778"

```

CDS

CDS

CDS

complement(join(144, 2404, 2450, .4378, 4501, .4660))
/note="ORF_ID:dd_02769"
/codon_start=1
/product="putative protein predicted proteins, Arabidopsis thaliana"
/protein_id="AA033192.1"
/db_xref="GI:20976603"
/translation="MEMNLRTFENRSKILILPIGDISPEKYEYSLIKTNITIELS
KILRSQNEJSEPEKISWDSVSLILNVEDISFORSEDEDTQTYKVFYIGVVDCKS
SLDQERKELEIVAAVAPSCVSSILCAFPDMQPDGLGALNIMPNNSDRHLTFY
LTLTLIDFSIMILKEEYKMYLSELDHYTSLTSQNYNNINNSGIIINSIGGN
NNNSLIGNSITIGANSNGVIGSSVYNSNNSLITTPDLSIKLMEIGAKKRGK
LNKRCDFPCILASGIPDAIKYVTLIDACSNQDMWMLASGCGYISAVLRKQSEF
TQASQTPRINSPNITSQNSNGSGNSGVITNPDSGNTNTNTNSSNN
NNNNNSNTNNNNNNNNNNNGVGNMTNMLIDQIOSDLKIRELASEAII
SYKKRVPOPEEDVYMLREAFYLVNTPREMEKINHLLIASKYOLEDELFNSRFL
LIDYSGKRFEPYVLEAFYLVNTPREMEKINHLLIASKYOLEDELFNSRFL
DEGSKSHSLFSLNOMKSNSTSGSRNNNNNNNNNNNNNNNNNNNNNNNNNN
AOVQNGNOSIOOVQSSNSRSSSSSRFPNNNIPELSYRKRFKGSVQYLYN
LISVSTNMDSLNISQOKOVEDLILHHSKALSIINTVHNHLLGLPIYAVSYLIQ
LPMOLPIYAVSYVLEAFERKSNNEFIFYSPGGSFNSGSGIGGGIGGGGSGSN
SNKKVVMAGGECSSVLELNSPESDFIPLQSTLTAGPEESYISGFIISLSDSM
LIDYISKPEPGLIKGVFRVAVNLSEHPVQGNSTISQYNELVRGDAKYDI
LEDSQPTLSSNNSSPSRNNNNNELPIYKIKAPKPIKADVPYGVKICVNLAGE
SRSELEFQNIQTQIDISVSLNEFDKALKKTKTLADNNOYFTDSDEFTSOWDSNI
INSLSDLPGETEYLRKTSKRFELGNQFIPLDYSNCEEESEEEFNTPNTNTNT
QRTLVGNQSTLTATTTATYVQYKRIPLDSTHGGVQVNPDIYVASSKIKSL
SPLVDRSNNSLFVEYDKDIEDCENYVDCLELFDICNDICSTPHFTISELE
LGIPVNDVNNKEINNNLSKFLIPSGITITIIYPMKRESLPOLPRAKQGYIKPK
KITIEVEYKCLIQYRKDLTSKIKLNNSSNTNGSLIDNIVITSKLIKHLKMDQ
VNGIYNNKOLVKTIEQGEETIEVLVNLNTPINGVLVHLPIIDQITNNNNNGK
NNGNIYVDSNGLVYIGSLTTPIDELPVGATFPHSLDVIFFERSEYFPIISCEIFTKQ
IIPASHSLVNVV"
join(5214, 6624, 6947, .7146, 8631, .8912, 9042, .9155,
9242, 9457, 9627, 10028)
/note="ORF_ID:dd_02770"
/codon_start=1
/product="Ribonuclease inhibitor."
/protein_id="AA033193.1"
/db_xref="GI:20976604"
/translation="MNVQVIMDEDOEIIICLNNDISLIEINDDDDEELISPPYK
QSIIRKILPIHMQIIFYTTLTLPGEFNSDEYVLAIKWDSKNDRSISLEAVCYWCSIT
IKASCRSLISLNNLEANTPIRTHYVLAKEIITNNINDSDVSDPDSDSDSDGDN
RKDYSNDLITYDYFDKRMKMYRRRRVYNDNSDKIGKCLATINLSPLNLSLNSM
TNIKEHYVAYIOSSLSNCNIRKILWADSCPIPDQGVYLSVLESCQIROLDISCG
SYNGLEAYGRALSNMOTTEKILMSVQNSYSPYCMSCNLNPHSQMSLKIGDIE
GMSALSDPTALINSGKELDGSNPGAGAITLASKNHSPLSIDLSNLSISTE
GFKDILLESSTPITNLNLSKRODLSPLANTVLSLAKFNTVLSNLSECNLOK
FQIISIALDSNNKRLKLNLSKRFNSNDLFIPLNTFKLNDVTLNENGIIDIDPSI
IHQLEITDLSLHEPGLTKNINLLREFRNLKLVNSPTEYVIAKANKQOQALASQ
ACVLAVDTLRLTINHGEGFGFGLKADAIADLERRRDVLALENEITPNKQALIE
DOKKALAEKKYKDKDQMODDILKLEAKTKKACKTTPVYLKQOITFELNKITKSE
QUNAKKRLVYMERKKAITVLSQNPQLEKTEISADTMSKFSRLNTHRLINQSQ
QLNLEMSHISAKOERTLVQIPQGGTGSNAVITVAPGTVITGGGGOATVEYESYDN
KDNQNYNYDGGGGGAGGSLQALARYADESAPSDLNKAGDIIISVIOQDDESGMTK
RAYEENEGIFPSTYEYV"
complement(join(10585, 10678, 10842, .10913, 11037, .11212))
/note="ORF_ID:dd_02772"
/codon_start=1
/product="hypothetical protein"
/protein_id="AA033194.1"
/db_xref="GI:20976605"
/translation="WNGVROKVVQDLPAGGPKRYARTSNPIPGAVIFAGVGSIM
AVGYTIFPSDYERNAAREDEKRRSLMLPIQLAENDINFLASPHQNYFTFRMPQPT
GKRAVALLDL"
complement(join(11580, .11786, 11891, .12133))
/note="ORF_ID:dd_02774"
/codon_start=1
/product="putative protein ubiquitin-conjugating enzyme,
Atcha pastoris"
/protein_id="AA033195.1"
/db_xref="GI:20976606"
/translation="AAALMEKEKYLQNEEEDLILYPRDESDLTRVAVLIKGPDPY

YENKGFELDLSVPTNYVLOPPTLKEFKVLRPHNPRTKGTGICIDILKTSMSAIVTLQS
VGRSIALISLSPEDSNPLCDAGLHNGVGRKHDLSARKYTRLGVC"
join(13214, .13969,14095, .14424)
/note="ORF_ID:dd_02776"
/codon_start=1
/product="ATG3609320./F3124_-19, 9./101"
/protein_id="AAM33196.1"
/db_xref="GI:20976607"
/translation="MVLINLNPSCIFIPILAMDEPDDDDDEQESLKNKKITTIPTINSQI
AIPILANNNDNNNSNNKTEQTTTTTITIKWTFIFNNKTSIFSGIIGRSIVLPLLVPI
YLILNLFAPETVYETTERKGIHSELSAMALVLOMLCNLYLCTDPGGSKPTISPY
YLILNFSSTSDNHKRWKCNCKNCHQPRHACHYCRNVCVLRMDHCOMQONCIGLFP
KYFVLFLEPTYSIIITVEPTLILKRSIELVRYKMETLPEFDLHLFLGLIITLII
AGISIMALMLOIALISKGLITLHSEIKRKRKQOQPVNLNKKYKDGSLINMSYVFG
NLSEFMLPPIPNMLKITSKKGDFIY"
-
complement(join(14699, .15027,15131, .15587))
/note="ORF_ID:dd_02779"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAM33197.1"
/db_xref="GI:20976608"
/translation="MKNDEPTELEHYHKDDDALEGEEOGEESLSDDDMNINGET
NDDDDDPADEEESSEDEKONOKYLRKKKRNKSSNGEFTKTIPTSLGKNTPIGI
LKNNSOPPKKRNRIWDEENLTINDMKSSYTMIDEPKPIHYHSESEETDGTAYLE
NKFLTELQALDOOKERSEMDSDDEDOOEKREKRNKSKKDKLIMRSDSDNDNDNE
DEDETEEKKENKFKFDNLRAKAYHFKVYRDLNMLSDDEQ"
complement(join(16148, .16663,16774, .17016,17117, .17599))
/note="ORF_ID:dd_02781"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAM33198.1"
/db_xref="GI:20976609"
/translation="MESVKNRIYKTOEHVHODHSYODGNHPRKLSKSEKSIIRSDYOL
LADSEMSDSDVODMOTKWEYLANITSPGGGLYIDNSRIHELSEGFHLOWMLPIL
KTKTIGIDNAELIISHSVKRIKONKONMORVADPHATKIDRYKRSKRNPTTLK
KKEVAERAKATITDYDSRKTOHDIRLEFSVDNLHPSROLFLDLSFTNHGIEVHVS
FRLSDQDPVAKELHSHRGLIDLOSLEPTKYRHSNMIKPDHNPVPHGSIIDEDQDA
FNSLDRSEQDTEIETPTTITTTTTYEVKPKRIDKLMDTESSPFRSLVSTNSQD
OKSYVASNTYINAGDLSTSTSTFTFENPQTPARKNIITPIIETSPASSSTYI
ISPRVSEYVYVNSKLPSE"
complement(join(18484, .20027,20118, .20424))
/note="ORF_ID:dd_03248"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAM33199.1"
/db_xref="GI:20976610"
/translation="MKSQNNINENONDKKKKSYVIRKFIYVYRYDAVORSEKQIENE
KELFESIPFNWILPSSLVOICLAGFYGMSIFPNPIDIYLRNDIKGAPITIT
MSITGILITPMPLENGPKGLGGLGIFVSPAHFAPRGVYTRIVLYIGTIGCS
GIGNSFTLSVSTLQKMFPRHKLPBGSVWGLNMGGITFSOCLTPIMLKRIYVITE
ITLGLGLITLLOALIFRLRPENPNTIINNKNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNINNTETNOEEBEENKENTIELKPPDNLKMTIIESITSYFRLIYIMF
ANGVCSNISERRLDFVODIFYFKDVRPASVAVTSVFNLGLIETLISEKIGRES
FNILITLIFELMLTITKILIESHPGFIISLIWYSPYGGGFGFVAPALITDEKGRNI
WSCHALITSNLNGVCGGFLTFPIYOTLENOHYVDNPDIYININFEVAVSCVIVGW
VCSPEIPKIVRYLPIPNLOLQFTFNPILPILINKKRPOLVLYENKEBEMNNTL
INIKSLALENNKNDNNNNNNNNNNNNNNNNNNNNNTPTSSSNSNNKLI"
21217, .23181
/note="ORF_ID:dd_02786"
/codon_start=1
/product="mitochondrial processing peptidase beta subunit
precursor (EC 3.4.24.64) (beta-MPP) (P=52)."
/protein_id="AAM33200.1"
/db_xref="GI:20976611"
/translation="MNIILFVPSIFEKKLVNWKIKILVANNQNSPIIVSNNTKSLIKN
HNSSEFTNANNNNNOVNLVONKSLISSSSYSGKNNNNNNKSLYTTSSNNNNKKE
ELVSTYPTPLNTLPLKYLGEKILSNQENNNNOKEKATISLPLGCTIVYSGOT
HEGVALGILYNAGTKYESPDODRGVFLNLEKMFETKNNNSSEIILKELETSNAMA
SSSRMINVSLSEVLRKQDEYVLSLSQIKSPYSSEBLREQDIEVCINRYEMITNSS
DQMLTELIMGVAEDDAGLGNLVATPEQYONITREKFLDALRYVGNRIVSTGAE
HSOVLDAVKDGIPEFKQDPESDISDTIYKGGTACVAGLTHKNHLSQLOFL
IEKQDLKQOQOQOQPOPSNDLNDNEBELNLNLEIQTKISIEQLEDAQVKESSI
IAFPGSGIYSAEKMDIINGIYQSLDGGSSSYGTGPGKGMQSRNLNLWVYSSHRVK

QDS	NCBAFLVFNKVSIFGSLITPQSGFLDGEHIVQLELMRRSSWQOELERARRSOKS QIULNLEMSRYQCDMDARHHLISFESTSRSPEDICLIDSVTLDDIKLISKLQASNPBV VSIANENEMRIALBAOYNQIVKQNSSTLFK Join(24601, .24764,24861, .25635) /note="ORF_ID:dd_03244" /codon_start=1 /product="hypothetical protein" /protein_id="AAH3201.1" /db_xref="GI:20976612"			
Query Match	4.3%;	Score 48.2;	DB 2;	Length 53150;
Best Local Similarity	43.3%;	Pred. No. 0.89;		
Matches 224;	Conservative 0;	Mismatches 293;	Indels 0;	Gaps 0;
OY	525	AGTTATCATATAATGACATCGCATCTTGGAAACCGTGTATACCTTTCTGCGGGCTTGTCCA	584	
DB	29376	AGATATATATATATATATATGAAATCAAAAGAAAGATATATTTGTGAAAGATGATGA	29435	
OY	585	TTCCGCTTTTGCATGTAGTGGAACGGGATACCTGTGAAATTTCAATTCGTGATGATG	644	
DB	29436	TAAATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	29495	
OY	645	TGCTCTGATATAATTTTGTCTAATATATTTGGAATATTCACACAGATTTAATGGCTGGCCA	704	
DB	29496	AGAATTTTAAAAACATCAAGTCAATAAATGATGAGAGAAAGAAAGATATTTACATTTATTTAA	29555	
OY	705	AGAACCTCAGCATATCAATATGCGGATCGATCCACGCTTTTGTATCAATCCAGATCAG	764	
DB	29556	TAAATCAATATGAAATGAAATTCAGATGATTCAGATGATTCAGATTCAGATTCAGATTC	29615	
OY	765	TATTTACCATTTAAAGAACCAATAGCGAATGTGTTTCGACACCAATGTTTCGAACCA	824	
DB	29616	TAAATATATAATATAATATAATATAATATAATATAATATAATATAATATAATATAAT	29675	
OY	825	ATTGGGACTGTTTAAACACAGTGGTGGCCGACGCAAAACCTGCTGACGTGGCCAACTTCG	884	
DB	29676	AGAGATTGCACACATTTTAACAATATAAACATATGAAATTAACGTTATGATGATATACCA	29735	
OY	885	TTTACTCAAGAAAAGATGTGCGACGCGAATAATTCATTTATGACCACTGATATATCA	944	
DB	29736	TCCATGACAGAAATTTTCCAAATTAAGATAATAGATTCAAAGTTTGAAGAAATTCCTTTAA	29795	
OY	945	CACCTTGAATTTAGCGATGATATCAACGCTTTGCCAGTGTATTTACGTCAACCGTGCACT	1004	
DB	29796	TAAATTTATGATGATGATGATTTTAAAGAACCATTCACATTCGTAATATGCAAGTTATACCA	29855	
OY	1005	TCTGCACATATATGACAACCTGTATATCTTGCTGCA	1041	
DB	29856	TCTTTTAAAGAAAGAAAGACGTCGTTGCAATTCACCA	29892	
RESULT 36				
AF067608	45713 bp	DNA	linear	INV 19-Apr-2002
LOCUS				
DEFINITION	Caenorhabditis elegans cosmid B0511, complete sequence.			
ACCESSION	AF067608			
VERSION	AF067608.1			
KEYWORDS	HTG.			
SOURCE	Caenorhabditis elegans.			
ORGANISM	Caenorhabditis elegans.			
REFERENCE	Eukaryote: Metazoa: Nematoda: Chromadorea: Rhabditiida; Rhabditioidea, Rhabditidae; Peloderinae; Caenorhabditis. 1 (bases 1 to 45713) Waterston, R.			
AUTHORS	Genome sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium			
JOURNAL	Science 282 (5396), 2012-2018 (1998)			
MEDLINE	98069613			
PubMed	9851916			
REFERENCE	2 (bases 1 to 45713) Tlin-Wollam, A., Suterer, C. and Ozersky, P.			
AUTHORS	The sequence of C. elegans cosmid B0511			
TITLE	Unpublished (2001)			
JOURNAL				

REFERENCE
AUTHORS
TITLE
JOURNAL
3 (bases 1 to 45713)
Waterston,R.
Direct Submission
Submitted (22-MAY-1998) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

REFERENCE
AUTHORS
TITLE
JOURNAL
4 (bases 1 to 45713)
Waterston,R.
Direct Submission
Submitted (30-AUG-2001) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA

REFERENCE
AUTHORS
TITLE
JOURNAL
5 (bases 1 to 45713)
Waterston,R.
Direct Submission
Submitted (27-SEP-2001) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA

REFERENCE
AUTHORS
TITLE
JOURNAL
6 (bases 1 to 45713)
Waterston,R.
Direct Submission
Submitted (05-OCT-2001) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA

REFERENCE
AUTHORS
TITLE
JOURNAL
7 (bases 1 to 45713)
Waterston,R.
Direct Submission
Submitted (19-APR-2002) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA

COMMENT
Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University
St. Louis, MO 63110, USA, and
Sanger Centre, Hinxton Hall
Cambridge CB10 1RO, England
email: twenemate@wustl.edu and jesse@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.
It may be shorter because we only sequence overlapping sections
once, or longer because we provide a small overlap between
neighboring submissions.

This sequence was finished as follows unless otherwise noted: all
regions were double stranded, sequenced with an alternate chemistry
or covered by high quality data (i.e., phred quality >= 30); an
attempt was made to resolve all sequencing problems, such as
compressions and repeats; all regions were covered by sequence from
more than one ml3 subclone.

For a graphical representation of this cosmid sequence and its
analysis see:
(www.wormbase.org/db/seq/sequence?name=B0511;class=Sequence)

NEIGHBORING COSMID INFORMATION

The 5' cosmid is C43H8, 2850 bp overlap; the 3' cosmid is C34B2,
300 bp overlap. Actual start of this cosmid is at base position
2847 of B0511; actual end is at 45613 of B0511.

NOTES:

Coding sequences below are the result of integration and manual
review of the following data: computer analysis using the program
GeneFinder (P. Green and L. Hillier, personal communication), the
large scale EST projects of Yuiji Kohara
(http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html) and The C.
elegans Genome cloning project (<http://wormdb.dicf.harvard.edu/>),
similarity to other proteins from Blastx analyses
(<http://blast.wustl.edu/>), sequence conservation with C. briggsae
using Jim Kent's WABA alignment program (Genome Research
10:1115-1125, 2000), individual C. elegans Genbank submissions,
and personal communications with C. elegans researchers. tRNAs
are predicted using the program tRNAscan-SE (Lowe, T.M. and

FEATURES
source
Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).
Location/Qualifiers
1..45713
/organism="Caenorhabditis elegans"
/strain="Bristol N2"
/db_xref="taxon:6239"
/chromosome="I"
/clone="B0511"
83..84
/note="S11 trans-splice acceptor; see yk290h4.5"
91..2919
/gene="B0511.6"
/note="for a graphical representation of this gene see:
(www.wormbase.org/db/seq/sequence?name=B0511.6;class=Sequence)"

CDS
join(91..402,451..1173,1677..1979,2239..2355,2475..2576,
2842..2819)
/gene="B0511.6"
/note="similar to DEAD-box helicases (Pfam: DEAD_hmm,
score: 262.89) (Pfam: hellicase_C_hmm, score: 76.21); coded
for by the following C. elegans cDNAs: cm01c4, yk2925.3,
yk9225.5, CESEF01F, CESEF227F, CESEF227R, yk290h4.5,
yk391e3.3, yk391e3.5, yk461a5.5, yk48201.3, yk48201.5,
yk483a8.5, yk727e5.3"
/codon_start=1
/product="Hypothetical protein B0511.6"
/protein_id="AAC17654.2"
/db_xref="GI:15375231"

gene
/translation="MGIDYKKVLRKRLKQKMDGHEKKAKVVEEPEDEEVE
EOKHOTESSTKVSSEFLTKTTFASLSEGVANLKAQVIGFTTWEIOAKSIDPL
LEGKDVLSAKTSGKTLAFLPALELHKLMKQNGTGVITVSPRELSTQVGL
SELGDSNLTGLVNGSNGRSKEDKLAKGVSIIIVATGRLDHLQNDNFVLRMKC
LIIDPADRLIDGFEITEMQOVLRLPKOROSMLFATSPKVDLTKALSNPVRS
VHEKAEKAEVEGIGOGYIYAPSDRLLLIFFELKTKKRVKVVSPSSNSYKFHEHL
NYIDPCKSIHKQKQKRTTTFPPQCAETGILCTVYARAGDIPAVNIDVQDPT
DPRETIHVGKRTANGTSGKALVLPREEIGFLRYLKAAVTLNFEFSFMSKANT
QSDENLISKNYLNKSAKAYKCYLRVDSISLSDIDFVYNDMLTAVSKSGFSSVP
FVLDLISNKPVKYIERSKLSCAGYRRKKOSFTFKAKK"
4800..6253
/gene="B0511.5"
/note="for a graphical representation of this gene see:
(www.wormbase.org/db/seq/sequence?name=B0511.5;class=Sequence)"

CDS
join(4800..4812,4910..5106,5571..5642,5691..5832,
6000..6253)
/gene="B0511.5"
/codon_start=1
/product="Hypothetical protein B0511.5"
/protein_id="AAC17653.1"
/db_xref="GI:15375231"

gene
/translation="MPTKGRVFLVLSGSDKDCVSEGTGRTTSITVPRDKGVETVQ
GKAGYTSVNVISFHDKFLTKVDRAVNICLVAPTGADVSLATVQPSLKIQVLA
AOPSCVEVEFDVRRRPALIVHVAPELHWCTDGTLDLFCMTVHDCVINEKSKR
RSKTIIDSGCSLIDTRLPLNLRKNSARVSKAFRRGDVAVVEFGVRLDIRNGT
SCPRPRC"
complement(6442..7843)
/gene="B0511.7"
/note="for a graphical representation of this gene see:
(www.wormbase.org/db/seq/sequence?name=B0511.7;class=Sequence)"

CDS
complement(join(6442..6918,7008..7178,7262..7434,
7486..7585,7641..7694,7739..7843))
/gene="B0511.7"
/note="similar to forkhead-associated (FHA) domains (Pfam:
FHA_hmm, score: 28.36); coded for by the following C.
elegans cDNAs: yk393e11.3, yk465a3.3, yk465a3.5,
yk626b10.5, yk718c8.3, yk718c8.5, yk727h2.5, yk728a2.3"
/codon_start=1
/product="Hypothetical protein B0511.7"
/protein_id="AAC17655.1"
/db_xref="GI:15375231"

CDS
/translation="WEIIDDSSKFRVSPWAKPNGAHLVDQGDALIKLIDDKKA
YFGRNNQVDFAVEHASCVRVALLLHYGLQRFALIDMDSHGTELGNNVRLPLEY

gene
/organism="Spiroplasma citri"
/strain="G113"
/db_xref="taxon:2133"
1076..2125
/gene="mreb2"
1076..2125
/gene="mreb2"
/note="actin-like protein"
/codon_start=1
/transl_table=4
/product="Mreb2"
/protein_id="AAL51057.1"
/db_xref="GI:17981533"
/translation="MANYKFEKGDYSEFLTLDGTANTVAVYSGGIVNEPSPMAYDTL
NNSLITPGNEAVKMGVCTHDIIRWATPLVDVSDMEAOOLKHFIRLKSGLMKR
SLVLATSGVEATELGERGLAKIAKDMGSHVLEEEVLAALAGINGLAGNVLIT
IGGTTDLAITSAGDVIYSKVAGKAOODEIOKIRVENVNLVIGPTAPQIKKEIT
ALXNIINEKTVRAGRDIIITSPREILIKPEIKNVILASFRITDLVYEDQPP
LAGDVINRGITICGGALIRGIVKIFESIFOLKVANDPLMCVIEGAKTYEKNGA
IERIQLDANEYKIS"
2385..3443
/gene="mreb3"
2385..3443
/gene="mreb3"
/note="actin-like protein"
/codon_start=1
/transl_table=4
/product="Mreb3"
/protein_id="AAL51058.1"
/db_xref="GI:17981534"
/translation="MAISDYAKNTFNSPKAPRKFIATDGTNSIAVVAAGRIFFN
ASVAVYEIGTKKTLALGNDAKKLIGKTHDKLEIYSPLENGAVTNLTFAEFTQIGKK
AKVADYKNAIVLTACPKNVTDLEKQATIKCKNIGADFYVIEEDSLAALGAGENIT
APKQFTLIDIGGKSSCAIISAGIVESKSIKTAGNYIDEILKYIRAKHTISGVN
AEQIKQIOGSLYKTKENKKMIIFGRDVTGMPKRAEVLDSERKLIIISFISITDLI
EVEKRTPAELAGDAVANGILVYGGCGKTAGIKEFLSNYFQIPVRVKNWETSVIDC
AYEKNRTDLIEEKKNR"
4424..5467
/gene="mreb4"
4424..5467
/gene="mreb4"
/note="actin-like protein"
/codon_start=1
/transl_table=4
/product="Mreb4"
/protein_id="AAL51059.1"
/db_xref="GI:17981535"
/translation="MALFNSAKKPAFVSMALGTANTLTVYVSGGIVNEPSPVAVRIH
ENRIIAGVGAEYKMIIGKNKSIIRIVRPVADGVINDIRATEQOLRIYFKLRTKOLK
SIMLAPSPVITTELEKALKIRIANNLGASKYVEEVEYKMAALGGVDVITKAGNVI
MGGTTDIAVWASGDIVIKSVKVAAGNTINDEAOKFIRSOYGLIETGTAQDIREIC
SLSKYPRERKKYVGRDVYSSLPREIELTPEEIRKRVKVPYSRIITDLVYQIEETPP
LAGGIFRNGITICGGALIKGIDIRFTDITDLPKISGEPILAVANGTKFESDIFD
LRFSMHVKEINLY"
5521..6579
/gene="mreb5"
5521..6579
/gene="mreb5"
/note="actin-like protein"
/codon_start=1
/transl_table=4
/product="Mreb5"
/protein_id="AAL51060.1"
/db_xref="GI:17981536"
/translation="MRPEPFPISLDIGTANVLAVYSGGVVNEPSPMAVNNKTNIS
IATCGAVDMYKTHGDIIRWATPLVDVIADEMAOOLKHFIRLKSMMKNTMKNAY
LACSGYTELEERALKVAKMGAEVYIIEEAKMAALGAGINIELPGHLLIDIGGC
TTDLAIISSGDIVVRSIKVAGNHEDDDIRIINSEVIAIAGQKTADEVKKEFIGSLV

Query Match	3.9%	Score 44.8	DB 1	Length 6942
Best Local Similarity	44.6%	Pred. No. 5.8		
Matches 175	Conservative 0	Mismatches 217	Indels 0	Gaps 0
OY	195	TGATGAAAGGTGGACGTCAGTTCAGTTCGCCGAATTTTCACCTTCATTTGATTCAGTCAGTATGTC	254	
Db	6039	TGTTTCAGATCAATTAAGTTGCCGGAATACACTTTGATGATGATATTCGTAATACAT	6098	
OY	255	GCGTACAGATCTCGATATCCACGCGTATTTTGTATCAACACACTGTCATTTGCTT	314	
Db	6099	TGCTGTACAAATATACATTTGATGGCAGAAAAACACCCAGATGTTTAAAAATTCAT	6158	
OY	315	TCATCATTAATTTTGTATCCAAAGTTTGATGTCGATATTCAGTACAAATCTTTACATGCA	374	
Db	6159	TGCTTCACTAGTATTAATACCAATACGAAAGATTCATGCAATCTACGACGATATTTGT	6218	
OY	375	AGCTGATTAACACGTTAGTCACAGATTGAGCTATCTAAATCACAATCGCTTTTCAAC	434	
Db	6219	TTACAGATTATCCAAAAGAGGAAAAATCTCTCAAGAAAAATTAAGAAATGTTTACTTAA	6278	
OY	435	TCAAAATTTCCCGATGCGCATGATGCCCTTATGAATTTTGGATGGTGACACACCGGTCA	494	
Db	6279	TGCCTTTCTTAATATTAACAGATTATGTAATTTGAATATATTAGAAAAATACACCAACGAAAT	6338	
OY	495	ACCACTTCAATTTGCTATCATCTATGTCAGCAGCTTTATCATAAATGACATCGATTCGA	554	
Db	6339	AGCTGCTGATATTATGAGAAATAGCAATTAACGTTGTGGGGAGGACGCTTAATATGAAA	6398	
OY	555	AACGCTGATACCTTCGCGGGGTGTCATT	586	
Db	6399	TATTCATTAATATTTCTTTGATATTTTCCAAAT	6430	
RESULT 39				
AL772373/c				
LOCUS		229420 bp	DNA	linear
DEFINITION	Danio rerio clone CH211-143A22, *** SEQUENCING IN PROGRESS ***			23
UNORDERED PLACES.				
AL772373				
AL772373.5	GI:21911694			
HTG: HTGS_PHASE1; HTGS_DRAFT.				
KEYWORDS				
SOURCE	Danio rerio.			
ORGANISM	Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.			
1 (bases 1 to 229420)				
Burton, J.				
Direct Submission				
Submitted (16-JUL-2002)	Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zface@sanger.ac.uk			
Clone requests: clonerequest@sanger.ac.uk				
On Jul 19, 2002 this sequence version replaced gi:21738685.				
Genome Center				
Center: Wellcome Trust Sanger Institute				
Center code: SC				
Web site: http://www.sanger.ac.uk				
Contact: zface@sanger.ac.uk				
Project Information				
Center project name: zc143A22				
Summary Statistics				
Assembly Program: XGAP4; version 4.5				
Chemistry: Dye-terminator; 100% of reads				
Consensus quality: 218244 bases at least 040				
Consensus quality: 221659 bases at least 030				
Consensus quality: 224352 bases at least 020				
Insert size: 227220; sum-of-contigs				

```

urres
source
1. location/Qualifiers
1. .229420
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone="CH211-143A32"
/clone_lib="CHORI-211"
1. .5835
misc_feature
/note="assembly_fragment:02197"
fragment_chain:1"
5936. .31562
misc_feature
/note="assembly_fragment:01245"
fragment_chain:1"
31663. .33883
misc_feature
/note="assembly_fragment:00196"
fragment_chain:1"
33984. .50157
misc_feature

```

		/note="assembly_fragment:00012
		fragment_chain:1"
		50258. .90171
		/note="assembly_fragment:00175
		fragment_chain:1"
		90272. .113829
		/note="assembly_fragment:03194
		fragment_chain:1"
		113930. .131772
		/note="assembly_fragment:00816
		fragment_chain:1"
		131873. .133959
		/note="assembly_fragment:01413
		fragment_chain:2"
		134060. .136653
		/note="assembly_fragment:00319
		fragment_chain:2"
		136754. .140568
		/note="assembly_fragment:02483
		fragment_chain:3"
		140669. .148865
		/note="assembly_fragment:02951
		fragment_chain:3"
		148966. .151193
		/note="assembly_fragment:00671"
		151294. .153798
		/note="assembly_fragment:01095"
		153899. .156506
		/note="assembly_fragment:01672"
		156607. .159528
		/note="assembly_fragment:01898"
		159629. .161685
		/note="assembly_fragment:01915"
		161786. .167182
		/note="assembly_fragment:02102"
		167283. .170032
		/note="assembly_fragment:02103"
		170133. .172193
		/note="assembly_fragment:02192"
		172294. .222532
		/note="assembly_fragment:02262"
		222633. .225124
		/note="assembly_fragment:02355"
		225225. .227236
		/note="assembly_fragment:02882"
		227337. .229420
		/note="assembly_fragment:03097"
BASE COUNT	73355 a	39719 c 40997 g 73136 t 2213 others
ORIGIN		
Query Match	3.8%, Score 44; DB 2; Length 229420;	
Best Local Similarity	48.8%; Pred. No. 8.4;	
Matches 119; Conservative 0; Mismatches 125; Indels 0; Gaps 0;		
QY	893 ACAAAAGATCTGCACGAGAATATCATTTGTAGCAACTGATATCACACCCTTG	952
Dd	94174 AGGAACAATCGCATTTCGGAGATTATCATTTCTTTCGGGGCATCCGAGAGCTCTGTC	94115
QY	953 AAATTAGGAGATATATCAACCTTTGCCAGTGTATTACGTACACCGTGCACTTCTGCAC	1012
Dd	94114 CATTTCGGGGATACTCCTAGCTTAGCAACTGATGTATTAAGGAGGACTCTCTCTAC	94055
QY	1013 ATAATGACACCGTATATCTTCCTCGACAGAACAAATGSAANTCGATCTACCACTTGG	1072
Dd	94054 TTATATATTAATTATGTTCTCAACCTCACTGCAATCAACACTTCATTGAGATACACGTTTTTA	93995
QY	1073 GCCTTCATGTTATATGGGTTTAAAGCATTCGATGTTCGCCGTCATATATACATT	1132
Dd	93994 TTCATATCTTAT	93935
QY	1133 CGTT 1136	
Dd	93934 ATTT 93931	

RESULT	40
AL645495/c	
LOCUS	AL645495
DEFINITION	81091 bp DNA linear PRI 28-NOV-2001
	Human DNA sequence from clone RP13-472L18 on chromosome 13,
	complete sequence.
ACCESSION	AL645495
VERSION	AL645495.4 GI:17154392
KEYWORDS	HTG.
SOURCE	human.
ORGANISM	Homo sapiens

REFERENCE 1 (bases 1 to 81091)

TITLE
JOURNAL

COMMENT

corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at

<http://www.sanger.ac.uk/Projects/C-elegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome 13, constructed by the Sanger Centre Chromosome 13 Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chr13>

RP13-472L18 is from the library RPc1-13.2 constructed by the group of Pieter de Jong. For further details see

<http://www.chori.org/bacpac/home.htm>

VECTOR: pBACE3.6

IMPORTANT: This sequence is not the entire insert of clone RP13-472L18. It may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true right end of clone RP13-472L18 is at 81091 in this sequence. The true right end of clone RP11-427O13 is at 2000 in this sequence.

FEATURES	Location/Qualifiers
source	1. .81091

BASE COUNT	26912	a	14138	c	14152	g	25889	t
ORIGIN								

Query Match	3.7%;	Score 43;	DB 9;	Length 81091;
Best Local Similarity	50.2%;	Pred. NO. 15;		
Matches 106;	Conservative 0;	Mismatches 105;	Indels 0;	Gaps 0;

[illegible]

Db 51844 GAGGAGACGAGATCATATGAATGCCCTATTTAAAGTTGAAATTAATCATCTGGTACACGCC 51785

QY 312 GTTTCATCCATTTTGTGAACAAAGTGGATCGGCATATGAGTAACTAGCTTTTACAT 371

Db 51784 TCACATTCATGTTTGGTACAGGCGCTAACGCGTTTCATTAAGCCTAAATATTATTATATGG 51725

QY 372 GGAAGCTGATTAATCAAGTATGACACAGATT 402

Db 51724 TTAAAGGACTTTAAGACTTTGCTGAGACGTT 51694

```
Search completed: March 28, 2003, 16:44:58
Job time : 4054 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2003, 13:31:58 ; Search time 304 Seconds

(without alignments)
8600.553 Million cell updates/sec

Title: US-10-054-562a-3

Perfect score: 1161

Sequence: 1 atgagatcttcgtctattgc.....ttcgtccaatcagaagcca 1161

Scoring table: IDENTITY_NDC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: N_Geneseq_101002.*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
25: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1161	100.0	1161	22	AAH24828	Nucleotide sequenc
2	1161	100.0	1161	22	AAH24829	Nucleotide sequenc
3	1161	100.0	1161	22	AAH24826	Nucleotide sequenc
4	1161	100.0	1779	22	AAH24827	Nucleotide sequenc
5	1161	100.0	1779	24	ABU57372	Dirofilaria immiti
6	504.2	43.4	1372	22	AAH24830	Nucleotide sequenc
7	504.2	43.4	1372	22	AAH24831	Nucleotide sequenc
8	504.2	43.4	1372	24	ABU57373	Dirofilaria immiti
9	461	39.7	892	22	AAH24839	Nucleotide sequenc

C	10	461	39.7	892	22	AAH24840	Nucleotide sequenc
C	11	461	39.7	892	24	ABU57374	Brugia malayi cuti
C	12	310.4	26.7	813	22	AAH24832	Nucleotide sequenc
C	13	310.4	26.7	813	22	AAH24833	Nucleotide sequenc
C	14	40	3.4	1664976	19	AAV21209	Methanococcus jann
C	15	39.4	3.4	9810	24	ABU32426	Human immune syste
C	16	38.8	3.3	14066	20	AAH99556	Nucleic acid sequenc
C	17	38	3.3	6799	24	ABK99779	CDNA encoding full
C	18	37.2	3.2	6242	24	ABO70961	Listeria monocytog
C	19	37	3.2	783	20	AAZ16979	Human gene express
C	20	37	3.2	783	20	AAH99054	Human gene express
C	21	37	3.2	795	20	AAZ17515	Human gene express
C	22	37	3.2	795	20	AAH98778	Human validated ca
C	23	37	3.2	1467	21	AACT9937	Human secreted pro
C	24	37	3.2	1467	22	AAH33034	Human colon cancer
C	25	37	3.2	1680	24	ABO70472	Listeria monocytog
C	26	36.8	3.2	6571	22	AAH89354	Arabidopsis thalia
C	27	36.6	3.2	1527	22	AAH5681	Nucleotide sequenc
C	28	36.6	3.2	8342	24	ABU32500	Human immune syste
C	29	36.4	3.1	1163020	24	ABO67197	Listeria innocua C
C	30	36.4	3.1	3011208	24	ABO69245	Listeria innocua D
C	31	36.2	3.1	510	22	ABA8451	Human breast cell
C	32	36.2	3.1	510	22	ABA8451	Human breast cell
C	33	36.2	3.1	510	22	ABA6355	Human foetal liver
C	34	36.2	3.1	510	22	ABA33417	Probe #11883 for g
C	35	36.2	3.1	510	22	AAK14772	Human brain expres
C	36	36.2	3.1	510	22	AAK40512	Human bone marrow
C	37	36.2	3.1	510	22	AAI65552	Probe #15232 used
C	38	36.2	3.1	1103	24	ABK77860	Bacillus clausii g
C	39	36.2	3.1	7810	22	AAH45434	Chemically pretrea
C	40	36.2	3.1	1481	22	AAH34068	Human colon cancer
C	41	36	3.1	1481	24	ABO54194	Human ovarian anti
C	42	35.8	3.1	44362	23	ABU29558	Drosophila melanog
C	43	35.6	3.1	541	18	AAH83869	DNA encoding a N-a
C	44	35.6	3.1	1195	24	ABO68674	Listeria monocytog
C	45	35.6	3.1	2365589	24	ABA90521	Genomic sequence o

ALIGNMENTS

RESULT 1	AAH24828	AAH24828 standard; DNA; 1161 BP.
ID	AAH24828	
XX	AAH24828;	
AC	22-AUG-2001 (first entry)	
DT		
XX		
DE	Nucleotide sequence of a cuticlin polypeptide.	
XX		
KW	Cuticlin; gene therapy; vaccine; helminth parasite; ss.	
XX		
OS	Dirofilaria immitis.	
XX		
FT	Key	Location/Qualifiers
FT	CDS	1..1161
FT		/*tag= a
FT		/product= "cuticlin"
FT		/note= "no termination codon given"
XX		
PN	US6248329-B1.	
XX		
PD	19-JUN-2001.	
XX		
PF	01-JUN-1999;	99US-0323427.
XX		
PR	01-JUN-1998;	98US-0087435.
XX		
PA	(CHAN/) CHANDRASHEKAR R.	
XX	(MORA/) MORALES T H.	
PI	Chandrashekar R, Morales TH;	

XX WPI: 2001-396953/42.
 DR P-PSDB: AAB84258.
 XX Novel nucleic acid molecule encoding a Dirofilaria immitis cuticlin
 PT protein is useful as a vaccine to prevent parasitic helminth infection
 PT
 XX
 XX
 PS Claim 1: Column 35-40; 29pp; English.
 CC The specification describes a Dirofilaria immitis nucleic acid molecule,
 CC that encodes a cuticlin protein. The Dirofilaria immitis nucleic acid
 CC molecule is useful as a probe to identify nucleic acid molecules, as a
 CC primer to produce nucleic acids, as a therapeutic reagent, e.g., gene
 CC therapy to inhibit cuticlin activity or production, or in a vaccine to
 CC prevent infection with helminth parasites. The cuticlin protein,
 CC antibodies raised against it, and inhibitory compounds of cuticlin may
 CC all be used in compositions to protect animals, especially mammals such
 CC as cats, dogs, and humans. The antibodies may be used to passively
 CC immunize an animal, or as reagents in assay to detect infection of
 CC helminths, or as tools to screen expression libraries to recover desired
 CC proteins. They may also be used to screen cytotoxic agents to the
 CC parasite and kill it directly. The present sequence encodes a cuticlin
 CC polypeptide.
 XX
 SQ Sequence 1161 BP; 338 A; 227 C; 235 G; 361 T; 0 other;
 Query Match 100.0%; Score 1161; DB 22; Length 1161;
 Best Local Similarity 100.0%; Pred No. 0; Mismatches 0; Indels 0; Gaps 0;
 Matches 1161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ATGATGATTCGCTTATTCCTTCTGATACACTTATTCATTCGCTTATTCGATTCG 60
 DB 1 ATGATGATTCGCTTATTCCTTCTGATACACTTATTCATTCGCTTATTCGATTCG 60
 QY 61 GTTGACAAATGCTGCAAGGTGAGCCAGAAATGGAATGAGACCACTTCAATCAATC 120
 DB 61 GTTGACAAATGCTGCAAGGTGAGCCAGAAATGGAATGAGACCACTTCAATCAATC 120
 QY 121 AATTTAATACAGTATGATTCGAAAGCATGTTATGTAAGGCTTTATGATCA 180
 DB 121 AATTTAATACAGTATGATTCGAAAGCATGTTATGTAAGGCTTTATGATCA 180
 QY 181 GAAGGTGCCGTAATGATGAGGTGAGCTCAAGTTGCCGAAATTCATTCATTTGAT 240
 DB 181 GAAGGTGCCGTAATGATGAGGTGAGCTCAAGTTGCCGAAATTCATTCATTTGAT 240
 QY 241 TCATGCAATGTTGGCGTACAGATCTGTAATCCACGCGTATTTTGTGAACAACA 300
 DB 241 TCATGCAATGTTGGCGTACAGATCTGTAATCCACGCGTATTTTGTGAACAACA 300
 QY 301 GTTGTCATTCGTTTCATTCATTAATTTGTACCAAGTTGATGTCATATGAGTACAA 360
 DB 301 GTTGTCATTCGTTTCATTCATTAATTTGTACCAAGTTGATGTCATATGAGTACAA 360
 QY 361 TCGTTTACATGAGAGCTGATAAACAAGTTAGTGCACAGATTTAGAGTTGACACA 420
 DB 361 TCGTTTACATGAGAGCTGATAAACAAGTTAGTGCACAGATTTAGAGTTGACACA 420
 QY 421 ACTGCTTTCAACCTCAAAATGTCGCGATGCCAGTATGCCGTTATTAATTTTGGATGG 480
 DB 421 ACTGCTTTCAACCTCAAAATGTCGCGATGCCAGTATGCCGTTATTAATTTTGGATGG 480
 QY 481 GGACCAACCGGTCACCACTTCATTTGCTATCATTTGGTCAGCAGTTATATCAATATG 540
 DB 481 GGACCAACCGGTCACCACTTCATTTGCTATCATTTGGTCAGCAGTTATATCAATATG 540
 QY 541 ACATGCGATTCTGAAACCGCTGATCTTCGCGCGTTGCCATTCCTGTTTGTGCAT 600
 DB 541 ACATGCGATTCTGAAACCGCTGATCTTCGCGCGTTGCCATTCCTGTTTGTGCAT 600
 QY 601 GATGGTAAGGATACCTGTGGAATTCGAATTCGATGAGATGCTCTTGTATAATAT 660
 DB 601 GATGGTAAGGATACCTGTGGAATTCGAATTCGATGAGATGCTCTTGTATAATAT 660

DB 601 GATGGTAAGGATACCTGTGGAATTCGAATTCGATGAGATGCTCTTGTATAATAT 660
 QY 661 TTGCTAAATATTTGGAAATTCACAGATTTAATGGCTGGCCCAAGAGCTCAGTATAC 720
 DB 661 TTGCTAAATATTTGGAAATTCACAGATTTAATGGCTGGCCCAAGAGCTCAGTATAC 720
 QY 721 AATATGCGGATGCATCACAGCTTTTCATCAATGCGAGATCAGTATTAACATAAAGAA 780
 DB 721 AATATGCGGATGCATCACAGCTTTTCATCAATGCGAGATCAGTATTAACATAAAGAA 780
 QY 781 CCAATATGCGAATGTGTGACCAACATGTTCAACACACAGAGATTCGAGCTGTAA 840
 DB 781 CCAATATGCGAATGTGTGACCAACATGTTCAACACACAGAGATTCGAGCTGTAA 840
 QY 841 ACAGGTGGTCCGACCAAAACCTGCGAGCTGGCACTTGTACTCAAGAAAGA 900
 DB 841 ACAGGTGGTCCGACCAAAACCTGCGAGCTGGCACTTGTACTCAAGAAAGA 900
 QY 901 TCTGCAACCGGAGATATCATTTGATGATGATGATGATGATGATGATGATGATGATG 960
 DB 901 TCTGCAACCGGAGATATCATTTGATGATGATGATGATGATGATGATGATGATGATG 960
 QY 961 GATGATATCAAGCTTTCGACGTTGATGATGATGATGATGATGATGATGATGATGAT 1020
 DB 961 GATGATATCAAGCTTTCGACGTTGATGATGATGATGATGATGATGATGATGATGAT 1020
 QY 1021 CAACCTGTAATCTGCTGACGATGACAAATGGAATGCAATGCAATGCAATGCAATGCA 1080
 DB 1021 CAACCTGTAATCTGCTGACGATGACAAATGGAATGCAATGCAATGCAATGCAATGCA 1080
 QY 1081 ATGTTATGAGGTTAAGCATTTGATGATGATGATGATGATGATGATGATGATGATG 1140
 DB 1081 ATGTTATGAGGTTAAGCATTTGATGATGATGATGATGATGATGATGATGATGATG 1140
 QY 1141 TTTGCTCCAAATCAGAAGCA 1161
 DB 1141 TTTGCTCCAAATCAGAAGCA 1161

RESULT 2
 AAH24829/c
 ID AAH24829 standard; DNA; 1161 BP.
 XX AAH24829;
 AC
 XX
 DT 22-AUG-2001 (first entry)
 XX
 DE Nucleotide sequence of antisense strand of cuticlin open reading frame.
 KW Cuticlin; gene therapy; vaccine; helminth parasite; ss.
 XX
 OS Dirofilaria immitis.
 PN US6248329-B1.
 XX
 PD 19-JUN-2001.
 XX
 PF 01-JUN-1999; 99US-0323427.
 XX
 PR 01-JUN-1998; 98US-0087435.
 XX
 PA (CHAN/) CHANDRASHEKAR R.
 PA (MORA/) MORALES T H.
 PI Chandrashekar R, Morales TH;
 XX
 DR WPI: 2001-396953/42.
 XX
 PT Novel nucleic acid molecule encoding a Dirofilaria immitis cuticlin
 PT protein is useful as a vaccine to prevent parasitic helminth infection
 PT
 PS Claim 1: Column 41-42; 29pp; English.

XX The specification describes a *Dirofilaria immitis* nucleic acid molecule,
CC that encodes a cuticlin protein. The *Dirofilaria immitis* nucleic acid
CC molecule is useful as a probe to identify nucleic acid molecules, as a
CC primer to produce nucleic acids, as a therapeutic reagent, e.g., gene
CC therapy to inhibit cuticlin activity or production, or in a vaccine to
CC prevent infection with helminth parasites. The cuticlin protein,
CC antibodies raised against it, and inhibitory compounds of cuticlin may
CC all be used in compositions to protect animals, especially mammals such
CC as cats, dogs, and humans. The antibodies may be used to passively
CC immunize an animal, or as reagents in assay to detect infection of
CC helminths, or as tools to screen expression libraries to recover desired
CC proteins. They may also be used to target cytotoxic agents to the
CC parasite and kill it directly. The present sequence is the complement
CC of an open reading frame encoding a cuticlin polypeptide.
XX
SO Sequence 1161 BP; 361 A; 235 C; 227 G; 338 T; 0 other:
Query Match 100.0%; Score 1161; DB 22; Length 1161;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AATATGATTCGCTTATGCTTCTGCTGACTACACTTATGCAATTCGTTATGATCCG 60
DB 1161 ATGATGATTCGCTTATGCTTCTGCTGACTACACTTATGCAATTCGTTATGATCCG 1102
QY 61 GTTGACAAATGCTGCAAGGTGAGCCGAAATGAATGTGACCAACTGCAATCAATC 120
DB 1101 GTTGACAAATGCTGCAAGGTGAGCCGAAATGAATGTGACCAACTGCAATCAATC 1042
QY 121 AATTTAATACAGTAATGCAATTCGAAAGCAATGTTATGTAAGGCTTTATGATCAA 180
DB 1041 AATTTAATACAGTAATGCAATTCGAAAGCAATGTTATGTAAGGCTTTATGATCAA 982
QY 181 GAAGGTTGCCGTAATGATGAAGGTGACGTCGAAGTCCGGAATTCACATTTGAT 240
DB 981 GAAGGTTGCCGTAATGATGAAGGTGACGTCGAAGTCCGGAATTCACATTTGAT 922
QY 241 TCATGAATGTGCGCGGTACAGATCAGTAATCCAGTGTGATTTTGAACAACAACT 300
DB 921 TCATGAATGTGCGCGGTACAGATCAGTAATCCAGTGTGATTTTGAACAACAACT 862
QY 301 GTTGCAATTCGTTCCATCCATATTTGTTACCAAAAGTTGATGTCATTCGATCA 360
DB 861 GTTGCAATTCGTTCCATCCATATTTGTTACCAAAAGTTGATGTCATTCGATCA 802
QY 361 TCGTTTACATGAGAGCTGATTAACAGTTAGTCACAGATTTGAGTATGAAATCACA 420
DB 801 TCGTTTACATGAGAGCTGATTAACAGTTAGTCACAGATTTGAGTATGAAATCACA 742
QY 421 ACGGCTTTTCAAACTCAAAATGTCCGAGTCCAGTATCCGTTATGAATTTGAGATG 480
DB 741 ACGGCTTTTCAAACTCAAAATGTCCGAGTCCAGTATCCGTTATGAATTTGAGATG 682
QY 481 GGAACCAACCGGTACACAGTTCAATTTGCTATGATTTGGTACGCAATTAATATG 540
DB 681 GGAACCAACCGGTACACAGTTCAATTTGCTATGATTTGGTACGCAATTAATATG 622
QY 541 ACATGCGATTCTGAAACCGTGTGATCTTCTGCGCGGTGTCATTCCTCTTGTGAT 600
DB 621 ACATGCGATTCTGAAACCGTGTGATCTTCTGCGCGGTGTCATTCCTCTTGTGAT 562
QY 601 GATGCTTACGCTGATCTGGAATTTCTAATGCTGATGATGCTCTTGAATATAT 660
DB 561 GATGCTTACGCTGATCTGGAATTTCTAATGCTGATGATGCTCTTGAATATAT 502
QY 661 TTGCTAAATATTTGGAATTTCAACAGATTTAATGGTCCCAAGAGCTACAGTATAC 720
DB 501 TTGCTAAATATTTGGAATTTCAACAGATTTAATGGTCCCAAGAGCTACAGTATAC 442
QY 721 AATATGCGGATGATCAGTATCTTCTATCAATGCCAGATAGTATACATTAAGAA 780
DB 441 AATATGCGGATGATCAGTATCTTCTATCAATGCCAGATAGTATACATTAAGAA 382

QY 781 CCAATATGCGAATGCTTTCGACCAATGTTCAAGAACCAAGGATTCGAGCTTTAAA 840
DB 381 CCAATATGCGAATGCTTTCGACCAATGTTCAAGAACCAAGGATTCGAGCTTTAAA 322
QY 841 ACAGGTGTCGCGGAGCAAAACCTGCTGACGTCGCAACTGCTTACTCAAGAAAAA 900
DB 321 ACAGGTGTCGCGGAGCAAAACCTGCTGACGTCGCAACTGCTTACTCAAGAAAAA 262
QY 901 TCTGCAAAACCGGAGATATCATTTGATGTCAGAACTATATCAACACCTTGAATATG 960
DB 261 TCTGCAAAACCGGAGATATCATTTGATGTCAGAACTATATCAACACCTTGAATATG 202
QY 961 GATGATATCAAGCTTTGCCAGTTGATTTAGTCACGTCACCTGCTTCAACATATGCA 1020
DB 201 GATGATATCAAGCTTTGCCAGTTGATTTAGTCACGTCACCTGCTTCAACATATGCA 142
QY 1021 CAACCTGTAATCTTGTGCAATCAAAATGAAATGCAATGTCACCAATTTGGCTCTCA 1080
DB 141 CAACCTGTAATCTTGTGCAATCAAAATGAAATGCAATGTCACCAATTTGGCTCTCA 82
QY 1081 ATGTTATGCTTAAACATTCGATTTGCTGCGCTCATTTATACATTTGCTTAAA 1140
DB 81 ATGTTATGCTTAAACATTCGATTTGCTGCGCTCATTTATACATTTGCTTAAA 22
QY 1141 TTTCGTCCAATCAGAAGCA 1161
DB 21 TTTCGTCCAATCAGAAGCA 1
RESULT 3
AAH24826
ID AAH24826 standard; DNA; 1779 BP.
AC AAH24826;
XX
XX 22-AUG-2001 (first entry)
DE Nucleotide sequence of sense strand of cuticlin gene.
XX
XX Cuticlin; gene therapy; vaccine; helminth parasite; ss.
KW
XX
XX *Dirofilaria immitis*.
OS
FH Key location/Qualifiers
FT CDS 167..1330
FT /*tag= a
FT /product= "cuticlin"
XX
PN US6248329-B1.
XX
PD 19-JUN-2001.
XX
PF 01-JUN-1999; 99US-0323427.
XX
PR 01-JUN-1998; 98US-0087435.
XX
PA (CHAN/) CHANDRASHEKAR R.
PA (MORA/) MORALES T H.
PI Chandrashekar R, Morales TH;
XX
XX WPI; 2001-396953/42.
DR P-PDB; AAB84258.
XX
XX Novel nucleic acid molecule encoding a *Dirofilaria immitis* cuticlin
PT protein is useful as a vaccine to prevent parasitic helminth infection
PT
XX
XX Claim 1; Column 33-34; 29pp; English.
XX
XX The specification describes a *Dirofilaria immitis* nucleic acid molecule,
CC that encodes a cuticlin protein. The *Dirofilaria immitis* nucleic acid

CC molecule is useful as a probe to identify nucleic acid molecules, as a
 CC primer to produce nucleic acids, as a therapeutic reagent, e.g., gene
 CC therapy to inhibit cuticlin activity or production, or in a vaccine to
 CC prevent infection with helminth parasites. The cuticlin protein,
 CC antibodies raised against it, and inhibitory compounds of cuticlin may
 CC all be used in compositions to protect animals, especially mammals such
 CC as cats, dogs, and humans. The antibodies may be used to passively
 CC immunize an animal, or as reagents in assay to detect infection of
 CC helminths, or as tools to screen expression libraries to recover desired
 CC proteins. They may also be used to target cytotoxic agents to the
 CC parasite and kill it directly. The present sequence encodes a cuticlin
 CC polypeptide.

XX Sequence 1779 BP; 575 A; 327 C; 302 G; 575 T; 0 other;

Query Match 100.0%; Score 1161; DB 22; Length 1779;

Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;

Matches 1161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGATGATTCGTCCTTATTCCTTCTGCTACACTTATGATTCGTCCTTATTCGATCCG 60
 DB 167 ATGATGATTCGTCCTTATTCCTTCTGCTACACTTATGATTCGTCCTTATTCGATCCG 226
 QY 61 GTTGACAATGGTGTGCAAGGTGAGCCAGAAATTGAATGTGAGCAACTTCATTAACATC 120
 DB 227 GTTGACAATGGTGTGCAAGGTGAGCCAGAAATTGAATGTGAGCAACTTCATTAACATC 286
 QY 121 AATTTAATACAGTATGATTCGCAAGGACATGTTATGTGAAGGCTTTATGATCA 180
 DB 287 AATTTAATACAGTATGATTCGCAAGGACATGTTATGTGAAGGCTTTATGATCA 346
 QY 181 GAAGGTGCGCTTAATGATGAAGGTGAGCTCAAGTTGCGGAATTCACCTTCATTGGAT 240
 DB 347 GAAGGTGCGCTTAATGATGAAGGTGAGCTCAAGTTGCGGAATTCACCTTCATTGGAT 406
 QY 241 TCATGAATGTGGCGGTACAGATCTCGAATCCACGCTGATTTTGTGAACAACT 300
 DB 407 TCATGAATGTGGCGGTACAGATCTCGAATCCACGCTGATTTTGTGAACAACT 466
 QY 301 GTTGATTCCTGTTTCATTCATTTATTTGTACCAAGTTGATGTCATATGAGATCA 360
 DB 467 GTTGATTCCTGTTTCATTCATTTATTTGTACCAAGTTGATGTCATATGAGATCA 526
 QY 361 TCGTTTACATGGAAGCTGATAAACAAGTTAGTGACAGATTGAGGATCTGAAATCACA 420
 DB 527 TCGTTTACATGGAAGCTGATAAACAAGTTAGTGACAGATTGAGGATCTGAAATCACA 586
 QY 421 ACTGCTTTCAACCTCAATTTGTCGCGATGCCAGTATGCCGTTATGAATTTTGGATGCT 480
 DB 587 ACTGCTTTCAACCTCAATTTGTCGCGATGCCAGTATGCCGTTATGAATTTTGGATGCT 646
 QY 481 GGACCAACCGGTCAACCAATTTGCTATCATTTGTCACGACGATTTATATATAATG 540
 DB 647 GGACCAACCGGTCAACCAATTTGCTATCATTTGTCACGACGATTTATATATAATG 706
 QY 541 ACATGCGATTTCGAAACCGTTGATCTTCTGCGCGGTGTCATTCCTGCTTTGTGAT 600
 DB 707 ACATGCGATTTCGAAACCGTTGATCTTCTGCGCGGTGTCATTCCTGCTTTGTGAT 766
 QY 601 GATGGTAACGGTATACGTGGAATTCGAATTCGATGAGATGTGCTTGTATTAATAT 660
 DB 767 GATGGTAACGGTATACGTGGAATTCGAATTCGATGAGATGTGCTTGTATTAATAT 826
 QY 661 TTGCTAAATTAATTTGGAATATCCACAGATTTAATGGCTGGCCAGAAAGCTACGATATC 720
 DB 827 TTGCTAAATTAATTTGGAATATCCACAGATTTAATGGCTGGCCAGAAAGCTACGATATC 886
 QY 721 AATATGCGGATGATCACAGCTTTTCTATCAATGCCAGATCAGTATTAACATTAAGAA 780
 DB 887 AATATGCGGATGATCACAGCTTTTCTATCAATGCCAGATCAGTATTAACATTAAGAA 946
 QY 781 CCAATATGCGGATGTGTTGACCACAATGTTAGAACCAACGAAGATTGGAGCTGTAA 840

DB 947 CCAATATGCGGATGTGTTGACCAACCAATGTTCAGAACCAAGATTGCGAGCTGTAA 1006
 QY 841 ACGGAGGTGGCCGACCAAAACCTGCTGAGCTGGCCAACTGCTTACTCAAGAAAGA 900
 DB 1007 ACGGAGGTGGCCGACCAAAACCTGCTGAGCTGGCCAACTGCTTACTCAAGAAAGA 1066
 QY 901 TCTGCAACCGGAGAAATATCATTTGATGTACCAAGTATGATACACCTTGAATTAAC 960
 DB 1067 TCTGCAACCGGAGAAATATCATTTGATGTACCAAGTATGATACACCTTGAATTAAC 1126
 QY 961 GATGATTAATCAAGCTTGGCCAGCTGATTAAGCTCACCGGTGCACTTCTGCAATATGA 1020
 DB 1127 GATGATTAATCAAGCTTGGCCAGCTGATTAAGCTCACCGGTGCACTTCTGCAATATGA 1186
 QY 1021 CAACCTGTAATCTTCTGCTGATACCAAAATGGAATCTGATGCACTTGGCTTCA 1080
 DB 1187 CAACCTGTAATCTTCTGCTGATACCAAAATGGAATCTGATGCACTTGGCTTCA 1246
 QY 1081 ATGTTTATGAGTTTAAAGCTTTCATTTGCTGCGGCTATTTACATTTGCTTTAA 1140
 DB 1247 ATGTTTATGAGTTTAAAGCTTTCATTTGCTGCGGCTATTTACATTTGCTTTAA 1306
 QY 1141 TTTTCGCAAAATCAGAGCA 1161
 DB 1307 TTTTCGCAAAATCAGAGCA 1327

RESULT 4
 AAH24827/c
 ID AAH24827 standard; DNA: 1779 BP.
 XX
 AC AAH24827;
 XX
 DT 22-AUG-2001 (first entry)
 XX
 DE Nucleotide sequence of antisense strand of cuticlin gene.
 XX
 KW Cuticlin; gene therapy; vaccine; helminth parasite; ss.
 XX
 OS Dirofilaria immitis.
 XX
 PN US6248329-B1.
 XX
 PD 19-JUN-2001.
 XX
 PE 01-JUN-1999; 99US-0323427.
 XX
 PR 01-JUN-1998; 98US-0087435.
 XX
 PA (CHAN/) CHANDRASHEKAR R.
 PA (MORA/) MORALES T H.
 PI Chandrashekar R, Morales TH;
 XX
 DR WPI: 2001-396953/42.
 XX
 PT Novel nucleic acid molecule encoding a Dirofilaria immitis cuticlin
 PT protein is useful as a vaccine to prevent parasitic helminth infection
 PT
 PS Claim 1; Column 35-36; 29pp; English.
 PS
 CC The specification describes a Dirofilaria immitis nucleic acid molecule,
 CC that encodes a cuticlin protein. The Dirofilaria immitis nucleic acid
 CC molecule is useful as a probe to identify nucleic acid molecules, as a
 CC primer to produce nucleic acids, as a therapeutic reagent, e.g., gene
 CC therapy to inhibit cuticlin activity or production, or in a vaccine to
 CC prevent infection with helminth parasites. The cuticlin protein,
 CC antibodies raised against it, and inhibitory compounds of cuticlin may
 CC all be used in compositions to protect animals, especially mammals such
 CC as cats, dogs, and humans. The antibodies may be used to passively
 CC immunize an animal, or as reagents in assay to detect infection of
 CC helminths, or as tools to screen expression libraries to recover desired

CC proteins. They may also be used to target cytotoxic agents to the
CC parasite and kill it directly. The present sequence represents the
CC complement of DNA encoding a cuticlin polypeptide.

SQ Sequence 1779 BP; 575 A; 302 C; 327 G; 575 T; 0 other;

Query Match	100.0%;	Score 1161;	DB 22;	Length 1779;
-------------	---------	-------------	--------	--------------

Best Local Similarity 100.08; Pred. No. 0;
Matches 1161; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

[illegible]

Accession	Sequence	Position
Oy	GATATATATCAAGCTTTGGCCAGTTGATTTACGACACCGGACACTTCTGCACATATATGGA	1020
Db	GATATATATCAAGCTTTGGCCAGTTGATTTACGACACCGGACACTTCTGCACATATATGGA	594
Oy	CAACCTTAATACTTGGTCGACGATACAAAATGSAATCTGCATCTACCACTTTGGCTTCCA	1080
Db	CAACCTTAATACTTGGTCGACGATACAAAATGSAATCTGCATCTACCACTTTGGCTTCCA	534
Oy	ATGTTTATGAGGTTTAAAGCATTTGCATTGATTCGTCCGTCATTATTACCATTTCTGTTTAAA	1140
Db	ATGTTTATGAGGTTTAAAGCATTTGCATTGATTCGTCCGTCATTATTACCATTTCTGTTTAAA	474
Oy	TTTGGTCCCAATCAGAAGGCA	1161
Db	TTTGGTCCCAATCAGAAGGCA	453

RESULT 5

ID ABL57372 standard; cDNA; 1779 BP

AC ABL57372;

DT 12-AUG-2002 (first entry)

DE *Dirofilaria immitis* cuticlin DiCut-1A cDNA

KW Cuticlin; helminth; parasite; DiCut-1A; vaccine; gene therapy;

KW anthelmintic; gene; ds.

OS *Dirofilaria immitis*.

FH	Key	Location/Qualifiers
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

ET :

```

1 /product 21000 in
2 /note= "the CDS is a
3 FT

```

XXI

052002037234-A1
FN
XX

FD 26-MAR-2002
XX

PE 20-MAR-2001;
XX

PR 01-JUN-1999; 99US-0323427.
XX

PA (CHAN/) CHANDRASHEKAR R.
PA (MOBA/) MOBATSE T II

XX
XX

XX

DR P-PSDB; ABB7

PT New parasiti

PT caused by pa

XX

XX

CC Dirofilaria

CC nucleic acid
CC sequence code

CC claimed: D1
CC expressed se

CC Ascaris and
CC encoded prot

cc invention pro
cc cuticlin pro

Claim 1, Page 18-19; 32pp; English.

The present sequence is the coding strand of cDNA encoding novel *Dirofilaria immitis* cuticlin Dicut-1A protein (see ABB/62/8). Nucleic acids comprising the complementary strand, the coding sequence coding strand and its complementary strand are also claimed. Dicut-1A cDNA was identified from a *D. immitis* larval expressed sequence tag DNA sequence on the basis of homology to *Ascaris* and *Caenorhabditis elegans* cut-1 cuticlin genes. The encoded protein has 91% identity with the *Ascaris* homologue. The invention provides *D. immitis* and *Brugia malayi* parasitic helminth cuticlin proteins, nucleic acids, antibodies, compounds that

CC inhibit cuticlin activity, and methods of obtaining them.
 CC Therapeutic compositions comprising such proteins, nucleic acids,
 CC antibodies and/or inhibitors, including genetic vaccines,
 CC recombinant virus vaccines and recombinant cell vaccines, are used
 CC to protect animals from diseases caused by parasitic helminths, and
 CC to inhibit the moulting of filarid larvae in an animal.

XX Sequence 1779 BP; 575 A; 327 C; 302 G; 575 T; 0 other;

Query Match 100.0%; Score 1161; DB 24; Length 1779;

Best Local Similarity 100.0%; Pred. No. 0; Matches 1161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGATGATTCGTTCTTATTCCTTCTGCTACTACATTTATGATTCGTTCTTATTCGATTCG 60
 DB 167 ATGATGATTCGTTCTTATTCCTTCTGCTACTACATTTATGATTCGTTCTTATTCGATTCG 226
 QY 61 GTTGACAAATGCTGTCAGAGTGAAGCAAGTGAATGATGAGACCACTTCAATACAAATC 120
 DB 227 GTTGACAAATGCTGTCAGAGTGAAGCAAGTGAATGATGAGACCACTTCAATACAAATC 286
 QY 121 AATTTAATACAGTATGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTC 180
 DB 287 AATTTAATACAGTATGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTC 346
 QY 181 GAAGTTGCGGTATGATGATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 240
 DB 347 GAAGTTGCGGTATGATGATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 406
 QY 241 TCATGCAATGTTGGCGGTGACAGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTC 300
 DB 407 TCATGCAATGTTGGCGGTGACAGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTC 466
 QY 301 GTTGATTCGTTCTTATTCCTTCTGCTACTACATTTATGATTCGTTCTTATTCGATTCG 360
 DB 467 GTTGATTCGTTCTTATTCCTTCTGCTACTACATTTATGATTCGTTCTTATTCGATTCG 526
 QY 361 TCGTTTACATGAGAGTGAATGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 420
 DB 527 TCGTTTACATGAGAGTGAATGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 586
 QY 421 ACTGCTTTCAACATCAATTTCCCGATGCGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 480
 DB 587 ACTGCTTTCAACATCAATTTCCCGATGCGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 646
 QY 481 GGACCAACCGGTCAACCACTTCAATTTGCTATCATGTCAGACAGTTCATATCAATAATGG 540
 DB 647 GGACCAACCGGTCAACCACTTCAATTTGCTATCATGTCAGACAGTTCATATCAATAATGG 706
 QY 541 ACATGCGATTCGAAACCGTGTATCTTCTGCGCGGTGTCATTCCTGCTTGTGAT 600
 DB 707 ACATGCGATTCGAAACCGTGTATCTTCTGCGCGGTGTCATTCCTGCTTGTGAT 766
 QY 601 GATGTAACGCTGATACGTGTGAATTTCTAATGCTGATGATGCTTCTTATTAATAT 660
 DB 767 GATGTAACGCTGATACGTGTGAATTTCTAATGCTGATGATGCTTCTTATTAATAT 826
 QY 661 TTGCTAAATATTTGGAATATCCACAGATTTAATGCTGCGCAAGAACTGACGATATC 720
 DB 827 TTGCTAAATATTTGGAATATCCACAGATTTAATGCTGCGCAAGAACTGACGATATC 886
 QY 721 AATATGCGGATGATACAGCTTTCTATCAATGCCAGATGATTAATACATTAAGAA 780
 DB 887 AATATGCGGATGATACAGCTTTCTATCAATGCCAGATGATTAATACATTAAGAA 946
 QY 781 CCAATATGCAATGCTTGTGACACCAATGTTGAGAACCAAGATTCGAGCTGTTAA 840
 DB 947 CCAATATGCAATGCTTGTGACACCAATGTTGAGAACCAAGATTCGAGCTGTTAA 1006
 QY 841 ACAGTGTGCGCGAGCAAAACCTGCTGAGCTGGCAAACTTCTTACTCAAGAAAGA 900
 DB 1007 ACAGTGTGCGCGAGCAAAACCTGCTGAGCTGGCAAACTTCTTACTCAAGAAAGA 1066

QY 901 TCTGCAGAACCGGAGATATCATTTGATGACGAACTGATATCAACACCTTGAAATTTAC 960
 DB 1067 TCTGCAGAACCGGAGATATCATTTGATGACGAACTGATATCAACACCTTGAAATTTAC 1126
 QY 961 GATGATATCAAGCTTTGGCAGTGTATTTACGTCACCCGCACTTCTGCAATATATGA 1020
 DB 1127 GATGATATCAAGCTTTGGCAGTGTATTTACGTCACCCGCACTTCTGCAATATATGA 1186
 QY 1021 CAACCTGATATCTGCTGCTAGTACAAATGGAATCTGATGTCACATTTGGCTTCTCA 1080
 DB 1187 CAACCTGATATCTGCTGCTAGTACAAATGGAATCTGATGTCACATTTGGCTTCTCA 1246
 QY 1081 ATGTTATGAGTGTTAAGCATTTGATTCGCTGCCGCTATTATTAACATTTGTTTAA 1140
 DB 1247 ATGTTATGAGTGTTAAGCATTTGATTCGCTGCCGCTATTATTAACATTTGTTTAA 1306
 QY 1141 TTTCTGCAAAATCAGAAAGCA 1161
 DB 1307 TTTCTGCAAAATCAGAAAGCA 1327

RESULT 6

AAH24830
 ID AAH24830 standard; DNA; 1372 BP.

AAH24830;

22-AUG-2001 (first entry)

Nucleotide sequence of sense strand of cuticlin gene.

Cuticlin; gene therapy; vaccine; helminth parasite; ss.

Dirofilaria immitis.

Key Location/Qualifiers

FT CDS 392..1207

FT /tag= a

FT /product= "cuticlin"

US6248329-B1.

19-JUN-2001.

01-JUN-1999; 99US-0323427.

01-JUN-1998; 98US-0087435.

(CHAN/) CHANDRASHEKAR R.

(MORA/) MORALES T H.

Chandrashekar R, Morales TH;

WPI: 2001-396953/42.

P-PSDB: AAB84259.

Novel nucleic acid molecule encoding a Dirofilaria immitis cuticlin

protein is useful as a vaccine to prevent parasitic helminth infection

Example 1; Column 43-44; 29pp; English.

The specification describes a Dirofilaria immitis nucleic acid molecule,

that encodes a cuticlin protein. The Dirofilaria immitis nucleic acid

molecule is useful as a probe to identify nucleic acid molecules, as a

primer to produce nucleic acids, as a therapeutic reagent, e.g., gene

therapy to inhibit cuticlin activity or production, or in a vaccine to

prevent infection with helminth parasites. The cuticlin protein,

antibodies raised against it, and inhibitory compounds of cuticlin may

all be used in compositions to protect animals, especially mammals such

as cats, dogs, and humans. The antibodies may be used to passively

immunize an animal, or as reagents in assay to detect infection of

helminths, or as tools to screen expression libraries to recover desired

CC proteins. They may also be used to target cytotoxic agents to the
CC parasite and kill it directly. The present sequence encodes a cuticlin
CC polypeptide.

XX Sequence 1372 BP; 456 A; 229 C; 281 G; 406 T; 0 other;

Query Match 43.4%; Score 504.2; DB 22; Length 1372;
Best Local Similarity 76.9%; Pred. No. 2.5e-131;
Matches 652; Conservative 0; Mismatches 193; Indels 3; Gaps 3;

```

QY 1 ATGATGATTCGCTTATTCGCTTCTCTGTAACATTTATG-CATTGCTTATTCGATTC 59
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 32 ATTATCATTTCTCTCTGTTATTCCTACTTTCTACTTACCTACGATCATCGATCCC 91
QY 60 GGTTCACATGTTGTCGACGTCGACGACGACGACGACGACGACGACGACGACGAC 119
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 92 TATTGACAAATGTCGACGACGACGACGACGACGACGACGACGACGACGACGACG 151
QY 120 CAATTTTATACACGATGATCGACGACGACGACGACGACGACGACGACGACGAC 179
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 152 CAATTTTCAATACGATGATCGACGACGACGACGACGACGACGACGACGACGAC 211
QY 180 AGAAGTTCGCTTATTCGCTTCTCTGTAACATTTATG-CATTGCTTATTCGATTC 239
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 212 GATGATGTCGCTTCAATGATGATGATGATGATGATGATGATGATGATGATGAT 271
QY 240 TTGATGATGTCGCTTCAATGATGATGATGATGATGATGATGATGATGATGATG 299
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 272 TTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 331
QY 300 TGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 359
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 332 AGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 391
QY 360 ATGCTTTTACATG-AAGCTATTAACACATGATGTC-ACAGATGATGATGATGATG 417
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 392 ATGCTTTTATGAAAGCTATTAACACATGATGTC-ACAGATGATGATGATGATG 451
QY 418 ACAAGCTTTTAACTCAATGTCGACGATGTCGACGATGTCGACGATGTCGACG 477
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 452 ACTACAGCATTCGAACTCAATGTCGACGATGTCGACGATGTCGACGATGTCGAC 511
QY 478 GGTGACCAACCGCTCAACGATTCATTCATTCATTCATTCATTCATTCATTCAT 537
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 512 GGTGACCAACCGCTCAACGATTCATTCATTCATTCATTCATTCATTCATTCAT 571
QY 538 TGGACATGCGATTCGAAACCGCTTATTCCTTGGCGGCTTGCATTCCTCTTGTG 597
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 572 TGGACATGCGATTCGAAACCGCTTATTCCTTGGCGGCTTGCATTCCTCTTGTG 631
QY 598 GATGATGCTTACGCTGATGATGATGATGATGATGATGATGATGATGATGATG 657
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 632 GATGATGCTTACGCTGATGATGATGATGATGATGATGATGATGATGATGATG 691
QY 658 TATTGCTTAAATTTGATGATTCGACGATTTATGCTTGGCCGACGACGATGAT 717
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 692 TATTGCTTAAATTTGATGATTCGACGATTTATGCTTGGCCGACGATGATGAT 751
QY 718 TACAAATATGCGATGATGATGATGATGATGATGATGATGATGATGATGATG 777
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 752 TATATATATGCGATGATGATGATGATGATGATGATGATGATGATGATGATG 811
QY 778 GAACCAATATGCGATGATGATGATGATGATGATGATGATGATGATGATGATG 837
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 812 GAACCAATATGCGATGATGATGATGATGATGATGATGATGATGATGATGATG 871
QY 838 AAAACAG 845
   ||| ||| |||
Db 872 AATTCGG 879

```

RESULT 7
AAH24831/c

```

ID AAH24831 standard; DNA; 1372 BP.
XX AC
XX AAH24831;
XX DT
XX 22-AUG-2001 (first entry)
XX DE
XX Nucleotide sequence of antisense strand of cuticlin gene.
XX KW
XX Cuticlin; gene therapy; vaccine; helminth parasite; ss.
XX OS
XX Dirofilaria immitis.
XX PN
XX US6248329-B1.
XX PD
XX .19-JUN-2001.
XX PF
XX 01-JUN-1999; 99US-0323427.
XX PR
XX 01-JUN-1998; 98US-0087435.
XX PA
XX (CHAN/) CHANDRASHEKAR R.
XX PA
XX (MORA/) MORALES T H.
XX PI
XX Chandrashekar R, Morales TH;
XX DR
XX WPI; 2001-396953/42.
XX PT
XX Novel nucleic acid molecule encoding a Dirofilaria immitis cuticlin
XX protein is useful as a vaccine to prevent parasitic helminth infection
XX
XX Example 1; Column 43-46; 29pp; English.
XX
XX The specification describes a Dirofilaria immitis nucleic acid molecule,
XX that encodes a cuticlin protein. The Dirofilaria immitis nucleic acid
XX molecule is useful as a probe to identify nucleic acid molecules, as a
XX primer to produce nucleic acids, as a therapeutic reagent, e.g., gene
XX therapy to inhibit cuticlin activity or production, or in a vaccine to
XX prevent infection with helminth parasites. The cuticlin protein,
XX antibodies raised against it, and inhibitory compounds of cuticlin may
XX all be used in compositions to protect animals, especially mammals such
XX as cats, dogs, and humans. The antibodies may be used to passively
XX immunize an animal, or as reagents in assay to detect infection of
XX helminths, or as tools to screen expression libraries to recover desired
XX proteins. They may also be used to target cytotoxic agents to the
XX parasite and kill it directly. The present sequence represents the
XX complement of DNA encoding a cuticlin polypeptide.
XX
XX Sequence 1372 BP; 406 A; 281 C; 229 G; 456 T; 0 other;
XX
Query Match 43.4%; Score 504.2; DB 22; Length 1372;
Best Local Similarity 76.9%; Pred. No. 2.5e-131;
Matches 652; Conservative 0; Mismatches 193; Indels 3; Gaps 3;
QY 1 ATGATGATTCGCTTATTCGCTTCTCTGTAACATTTATG-CATTGCTTATTCGATTC 59
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1341 ATTATCATTTCTCTCTGTTATTCCTACTTTCTACTTACCTACGATCATCGATCCC 1282
QY 60 GGTTCACATGTTGTCGACGTCGACGACGACGACGACGACGACGACGACGACGAC 119
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1281 TATTGACAAATGTCGACGACGACGACGACGACGACGACGACGACGACGACGACG 1222
QY 120 CAATTTTATACACGATGATCGACGACGACGACGACGACGACGACGACGACGAC 179
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1221 CAATTTTCAATACGATGATCGACGACGACGACGACGACGACGACGACGACGAC 1162
QY 180 AGAAGTTCGCTTATTCGCTTCTCTGTAACATTTATG-CATTGCTTATTCGATTC 239
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1161 GATGATGTCGCTTCAATGATGATGATGATGATGATGATGATGATGATGATGAT 1102
QY 240 TTGATGATGTCGCTTCAATGATGATGATGATGATGATGATGATGATGATGATG 299
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1101 TTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1042

```

QY	300	TTTTTTCATTTTCGTTTCCATTTATTTGGTATACCAAAATGATGCTGATTCGACATCA	359
Db	1041	AGTTGTCAATTCATTTTCATTCCAAATTTGCTTACCAAAAATAGATGACGATATCTGATACA <td>982</td>	982
QY	360	ATGCTTTTACATG- /AGCTGATATAAACAGTTAGTGC- /ACAGATTGAGGTATCTGAATC <td>4171</td>	4171
Db	981	ATGTTTTATATATGCAAACTGATATAGACCGTTAGTACTGTCCTTTGAAGTATCTGAATG <td>9222</td>	9222
QY	418	ACAATCTGCTTTTCAAACTCAAAATTTGTCCCATGTCAGATATGCCGTTATGAAATTTTGAT <td>4777</td>	4777
Db	921	ACTACAGATTCACAAACACAAAGTGGTACCAATGGCCGATATGCGATATGAGATTTTGAA <td>8628</td>	8628
QY	478	GGTGGACCAACCGGTCAACAGTTCATTTGCTATTCATTTGCTGCAGCCAGTTTATCATAA <td>5377</td>	5377
Db	861	GGTGGACCAACCGGTGCACCTGTTTCGATTTGCTATTCATTTGCTGCAGCTTATCATAA <td>8022</td>	8022
QY	538	TGGACATCCGATTTCTGAAACCGTGTGATCTTTCTGCGCGGTTGTCATTCCTGCTTGTG <td>5977</td>	5977
Db	801	TGGACATGTTGATTTTCAGAGACTACAGATTCATTTCTGCTGATTTAGATTCATGTTGTG <td>7422</td>	7422
QY	598	GATGATGGTAAACGGGATACCTGTGGAAATTTCTAAATGCTGATGGATGTGCTTTGATAA <td>6577</td>	6577
Db	741	GATGATGGAAAAAGGTGATGCGAGTGGAGTTGTGATGAAGAAGGATGTGCTTTGGACAA <td>6822</td>	6822
QY	658	TATTTGCTAAATTAATTTGGAATATCCAAAGATTTAAATGGCTGGCCAAAGAGCTCACGTA <td>7171</td>	7171
Db	681	TATTTACTCAATTAATTTTGATATATATATACAGATTTAATGCTGGCCAAAGAGCTCATGTT <td>6222</td>	6222
QY	718	TACAAATATGCGGATCGATCACAGCTTTTCTATCATGTCAGATCACTATTTACCATTTAA <td>7777</td>	7777
Db	621	TATTAATATGAGATTCGATCGATCAGACTTACTATCATTAATGCCAGATTTAGTATACAAATTA <td>5622</td>	5622
QY	778	GAACCAATTACGGAATGTTTCGACCCACAATGTTCAAAACACACAGAGATTCGAGCTGTT <td>8377</td>	8377
Db	561	GAGCCACATTCAGGATGTCTTCGACCAACAATGACACAGACGCAAGGATTTGGTGCATA <td>5022</td>	5022
QY	838	AAAACAGG 845 <td></td>	
Db	501	AAATCTGG 494 <td></td>	
RESULT 8			
ABL57373			
ID	ABL57373	standard; cDNA; 1372 BP.	
XX	AC	ABL57373;	
XX	DT	12-AUG-2002 (first entry)	
XX	DE	Dirofilaria immitis cuticlin DiCut-1B cDNA.	
XX	KW	Cuticlin: helminth; parasite; DiCut-1B; vaccine; gene therapy;	
XX	KW	antelmintic; gene; ds.	
XX	OS	Dirofilaria immitis.	
XX	FT	Key	Location/Qualifiers
XX	FT	CDS	392..1207
XX	FT	/*tag- a	/product= "DiCut-1B"
XX	FT	/note= "the CDS is also specifically claimed in	Claim 1"
XX	PN	US2002037294-A1.	
XX	PD	28-MAR-2002.	
XX	PF	20-MAR-2001; 2001US-0812642.	
XX	PR	01-JUN-1999; 99US-0323427.	
XX	DA	(CHAN/) CHANDRASHEKAR R.	

Query Match	Best Local Similarity	Matches 652: Conservative	Score 504.2; DB 24; Length 1372: 76.9%; Pred. No. 2.5e-131;	Mismatches 193;	Indels 3;	Gaps 3;
Qy 1	ATGATGATTGCGCTATATGCTTCTTGCTGATACACTTATG-CATGTGCTATTGCATGCC	59				
Db 32	ATTATCATATTATTCCTCTGTTATTCCTACTTTCTTACTTGACCTACGTAATCGATGCC	91				
Qy 60	GTTTGACAAATGGTGTGCAAGGTGAGCAGCAAAATTAATGTGGACCACTTCAATAACAT	119				
Db 92	TATTGCAATGGTGTGCAAGGTGAGCAGCAAAATTAATGTGGACCACTTCAATAACAT	151				
Qy 120	CAATTTTAATACACGTAATGCAATTCGAAGAGCATTTTATGTGAAGGCTTTATGATCA	179				
Db 152	CAATTTTAATACGTAATGCAATTTGGAAGGAGCAGCATATGTAAGGAGCTTATGATCA	211				
Qy 180	AGAAGCTGCGGTATGATGAAAGGTGGAGCGCAATTTGCCGAATTTCACTTCCATTTGA	239				
Db 212	GGATGAAATGTGCTGACATAGTAATGAGCAGGAGTACTGGAATGCAATTTGGCAATGA	271				
Qy 240	TTTCATGCAATGTGCGGTGACGATCTCTGTAATCCACAGTGGTATTTTGTAAACAAC	299				
Db 272	TTTCATGCAATGTGCGGTGACGATCTCTGTAATCCACAGTGGTATTTTGTAAACAAC	331				
Qy 300	TGTTGTGCAATGTGCGGTGACGATCTCTGTAATCCACAGTGGTATTTTGTAAACAAC	359				
Db 332	AGTTGTGCAATGTGCGGTGACGATCTCTGTAATCCACAGTGGTATTTTGTAAACAAC	391				
Qy 360	ATGCTTTTAATGAGG-AAAGCTGATTAAGAGTGAAGG-ACAGATGAGCATTTGCAATTC	417				
Db 392	ATGCTTTTAATGAGG-AAAGCTGATTAAGAGTGAAGG-ACAGATGAGCATTTGCAATTC	451				
Qy 418	ACAATGCTTTTCAAACTCAAAATTTGCCAGTGGCAGTATCCGTTATGAATTTTGGAT	477				
Db 452	ACTGACAGATTCGAAAGCTCAAGTGGTACCAATGCGCGGATGATGAGATTTTGGAA	511				
Qy 478	GGTGGACCAACCGGTGCAACGATTTGCTATCATTTGCTGACGCCAGTTATCATPAA	537				
Db 512	GGTGGACCAACCGGTGCAACGATTTGCTATCATTTGCTGACGCCAGTTATCATPAA	571				

```
QY 538 TGGACATGGGATTCGTAACCGTTGATACCTTCGCGGGTGGCCATCTCCTGCTGTC 597
DB 572 TGGACATGGGATTCGTAACCGTTGATACCTTCGCGGGTGGCCATCTCCTGCTGTC 631
QY 598 GATGATGTAACGGTGTACTGTGGAATTTCTAAATGCTGATGATGCTGCTGTGATAAA 657
DB 632 GATGATGTAACGGTGTACTGTGGAATTTCTAAATGCTGATGATGCTGCTGTGATAAA 691
QY 658 TATTTGCTAAATATTTTGGAAATTCACAGATTTAAATGCTGGCCAGAAAGCTCACGTA 717
DB 692 TATTTGCTAAATATTTTGGAAATTTAAATGCTGGCCAGAAAGCTCACGTA 751
QY 718 TACAAATATGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 777
DB 752 TATTAATATGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 811
QY 778 GAACCAATATGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 837
DB 812 GAGCAGATGATGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 871
QY 838 AAAACAGG 845
DB 872 AAATCTGG 879

RESULT 9
AAH24839
ID AAH24839 standard; DNA: 892 BP.
AC AAH24839;
XX
XX 22-AUG-2001 (first entry)
DE Nucleotide sequence of a cuticlin polypeptide.
XX
XX Cuticlin: gene therapy; vaccine; helminth parasite; ss.
XX
XX Brugia malayi.
XX
XX Key Location/Qualifiers
XX CDS 158..892
XX FT /*tag= a
XX FT /product= "cuticlin"
XX FT /note= "no termination codon given"
XX
XX US6248329-B1.
XX
XX 19-JUN-2001.
XX
XX 01-JUN-1999; 9905-0323427.
XX
XX 01-JUN-1998; 9805-0087435.
XX
XX (CHAN/) CHANDRASHEKAR R.
XX (MORA/) MORALES T H.
XX
XX Chandrashekar R, Morales TH;
XX
XX WPI; 2001-396953/42.
XX
XX P-PSDB; AAB84260.
XX
XX Novel nucleic acid molecule encoding a Dirofilaria immitis cuticlin
XX protein is useful as a vaccine to prevent parasitic helminth infection
XX
XX
XX Disclosure; Column 51-54; 29pp; English.
XX
XX The specification describes a Dirofilaria immitis nucleic acid molecule,
XX that encodes a cuticlin protein. The Dirofilaria immitis nucleic acid
XX molecule is useful as a probe to identify nucleic acid molecules, as a
XX primer to produce nucleic acids, as a therapeutic reagent, e.g., gene
XX therapy to inhibit cuticlin activity or production, or in a vaccine to
XX prevent infection with helminth parasites. The cuticlin protein,
```

```
CC antibodies raised against it, and inhibitory compounds of cuticlin may
CC all be used in compositions to protect animals, especially mammals such
CC as cats, dogs, and humans. The antibodies may be used to passively
CC immunize an animal, or as reagents in assay to detect infection of
CC helminths, or as tools to screen expression libraries to recover desired
CC proteins. They may also be used to target cytotoxic agents to the
CC parasite and kill it directly. The present sequence encodes a cuticlin
CC polypeptide.
XX
XX SQ Sequence 892 BP; 290 A; 142 C; 168 G; 292 T; 0 other;
XX
XX Query Match 39.7%; Score 461; DB 22; Length 892;
XX Best Local Similarity 78.1%; Pred. No. 2.8e-119;
XX Matches 554; Conservative 0; Mismatches 155; Indels 0; Gaps 0;
XX
QY 21 TTTCTGATACATCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 80
DB 184 TTTGTCATATATGATATATGATATGATATGATATGATATGATATGATATGATATGATATGAT 243
QY 81 TGAGCCGAATTTGAATGTGACCACTTCAATATCAATATCAATATTTATATACGTAATGC 140
DB 244 TGAACCTGAATTTGAATGTGACCACTTCAATATCAATATTTATATCTCGAAATTC 303
QY 141 ATTGGAAGGACATGTTATGTAAGGATTTATGATCAAGAGCTTGCCTAATGATGA 200
DB 304 TTTTGAAGGACATGTTATGTAAGGATTTATGATCAAGAGCTTGCCTAATGATGA 363
QY 201 AGGTGAGCGTCAAGTTCGCGGAAATTTGATTCATTTGATTCATGATGATGTCGCTAC 260
DB 364 AGGTGAGCGTCAAGTTCGCGGAAATTTGATTCATTTGATTCATGATGATGTCGCTAC 423
QY 261 ACGATCTGATATCCAGTGTGATTTTGTATCAACAACATGTCATTTGCTTCAATCC 320
DB 424 ACGTTCGTTAAATCCAGTGTGATTTTGTATCAACAACATGTCATTTGCTTCAATCC 483
QY 321 ATTATTTGTTACCAAGTGTGATTCGTCATATGATGATGATGATGATGATGATGATGATGAT 380
DB 484 ACGATTTATCAAAAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 543
QY 381 TAAACAGTTAGTGCACAGATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 440
DB 544 TAAACAGTTAGTGCACAGATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 603
QY 441 TGTCCGATGCGCATATGCGCTTATGAAATTTGATGATGATGATGATGATGATGATGATGAT 500
DB 604 GTTACCAATGCGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 663
QY 501 TCAATTTGCTATCATTTGTCAGCCAGTTTATCATTAATGATGATGATGATGATGATGATGAT 560
DB 664 CCAATATGCTATATTTGACACACCGTTTATCATTAATGATGATGATGATGATGATGATGAT 723
QY 561 TGATACCTTCTGCGCGGTGTGTCATTCCTGCTTGTGATGATGATGATGATGATGATGATGAT 620
DB 724 TGATACCTTCTGCGCGGTGTGTCATTCCTGCTTGTGATGATGATGATGATGATGATGATGAT 783
QY 621 GGAATTTCTAAATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 680
DB 784 TAAATTAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 843
QY 681 TCCACAGATTTAATGCTGGCCAGAAAGCTCAGTATACAAATATGCG 729
DB 844 TCCACAGATTTAATGCTGGCCAGAAAGCTCAGTATACAAATATGCG 892

RESULT 10
AAH24840/c
ID AAH24840 standard; DNA: 892 BP.
AC AAH24840;
XX
XX 22-AUG-2001 (first entry)
XX
XX Nucleotide sequence of antisense strand of cuticlin open reading frame.
```

XX Cuticlin; gene therapy; vaccine; helminth parasite; ss.
 XX Brugia malayi.
 OS US6248329-B1.
 XX 19-JUN-2001.
 PD 01-JUN-1999; 99US-0323427.
 XX 01-JUN-1998; 98US-0087435.
 XX (CHAN/) CHANDRASHEKAR R.
 XX (MORA/) MORALES T H.
 XX Chandrashekar R, Morales TH;
 DR WPI: 2001-396953/42.
 XX
 PT Novel nucleic acid molecule encoding a Dirofilaria immitis cuticlin
 PT protein is useful as a vaccine to prevent parasitic helminth infection
 PS
 XX Disclosure: Column 55-56; 29pp; English.
 XX
 CC The specification describes a Dirofilaria immitis nucleic acid molecule,
 CC that encodes a cuticlin protein. The Dirofilaria immitis nucleic acid
 CC molecule is useful as a probe to identify nucleic acid molecules, as a
 CC primer to produce nucleic acids, as a therapeutic reagent, e.g., gene
 CC therapy to inhibit cuticlin activity or production, or in a vaccine to
 CC prevent infection with helminth parasites. The cuticlin protein,
 CC all be used in compositions to protect animals, especially mammals such
 CC as cats, dogs, and humans. The antibodies may be used to passively
 CC immunize an animal, or as reagents in assay to detect infection of
 CC helminths, or as tools to screen expression libraries to recover desired
 CC proteins. They may also be used to target cytotoxic agents to the
 CC parasite and kill it directly. The present sequence is the complement
 CC of an open reading frame encoding a cuticlin polypeptide.
 CC
 SQ Sequence 892 BP; 292 A; 168 C; 142 G; 290 T; 0 other;
 Query Match 39.7%; Score 461; DB 22; Length 892;
 Best Local Similarity 78.1%; Pred. No. 2.8e-119;
 Matches 554; Conservative 0; Mismatches 155; Indels 0; Gaps 0;

DB 349 TAAGACTGTTAGCACAAATTTGAAGTTCCGAATAGACACCGTATTTGTACACAAAT 290
 QY 441 TGTCCGATGCCAGTATGCGGTATGAAATTTTGATGGTGACCAACCGGTCAACCACT 500
 DB 289 GGTACCAATGCCCTGTGTGTAGATATGATTCGTGATGGTGTCCAAACCGGACACTCT 230
 QY 501 TCAATTTGCTATTCATGCTGACACCAAGTTATCATTAATGAGACATGGCATTTGAAACCT 560
 DB 229 CCAGTATGCTAATATGACAAACCGGTATATCATTAATGACATGATTTGTGAAACAGT 170
 QY 561 TGATCTTTCTGCGCGGTGTCCATTCCTGCTTTTGATGATGATGATGATGATGATGAT 620
 DB 169 TGATACCTTCTGCTTTGTGATGATGATGATGATGATGATGATGATGATGATGAT 110
 QY 621 GGAATTTCTAATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 680
 DB 109 TAATTTAATTAATGAAGAAGATGTCCATTAATGATGATGATGATGATGATGATGAT 50
 QY 681 TCCAAACAGATTTAATGCTGCGCCAGAACCTCAGCATATCAATATGCG 729
 DB 49 TCCAACTGATCTAATGCTGCGCCAGAACCTCAGCATATCAATATGCG 1
 RESULT 11
 ABL57374
 ID ABL57374 standard; cDNA; 892 BP.
 XX
 AC ABL57374;
 XX
 DT 12-AUG-2002 (first entry)
 XX
 DE Brugia malayi cuticlin BmCut-1A cDNA.
 XX
 KW Cuticlin; helminth; parasite; BmCut-1A; vaccine; gene therapy;
 KW anthelmintic; gene; ds.
 OS Brugia malayi.
 OS
 FH
 FH Key Location/Qualifiers
 FT 158..892
 FT CDS
 FT /*tag= a
 FT /product= "BmCut-1A"
 FT /partial
 FT /note= "the CDS does not include a stop codon"
 XX
 PD US2002037294-A1.
 XX
 PD 28-MAR-2002.
 XX
 PF 20-MAR-2001; 2001US-0812642.
 XX
 PR 01-JUN-1999; 99US-0323427.
 XX
 PA (CHAN/) CHANDRASHEKAR R.
 PA (MORA/) MORALES T H.
 XX
 PI Chandrashekar R, Morales TH;
 PI
 DR WPI: 2002-401306/43.
 DR P-PSDB: ABB76280.
 XX
 PT New parasitic helminth cuticlin proteins and nucleic acids isolated
 PT form Dirofilaria immitis, useful for protecting animals from diseases
 PT caused by parasitic helminths or for inhibiting molting of filarid
 PT larvae in an animal -
 XX
 PS Claim 9; Page 27-28; 32pp; English.
 XX
 CC The present sequence is the coding strand of a partial cDNA
 CC clone for novel Brugia malayi BmCut-1B protein (see ABB76280).
 CC The complementary strand of this sequence is also claimed. The
 CC BmCut-1A cDNA was obtained from adult female cDNA by PCR
 CC amplification using primers based on Dirofilaria immitis cut-1

Db 254 GTGATGCTGAGTTCGATGTAAGAGATGCTTTGGACAAATATTACTCAATA 313
 QY 671 ATTGGAATATCCAGATTTTATGCTGGCCAGAGAGCTCAGTATCAAAATATGCG 730
 Db 314 ATTGGAATATATATACAGATTTTATGCTGGCCAGAGAGCTCAGTATCAAAATATGCG 373
 QY 731 ATGATCAGACGCTTTTCTATCAATGCGCAGATCAGTATTAACATTAAGAACAATATGCG 790
 Db 374 ATGATCAGACGCTTTTCTATCAATGCGCAGATCAGTATTAACATTAAGAACAATATGCG 433
 QY 791 AATGTGTGACACATTTTCAAGACCAAGGATTCG----- 830
 Db 434 AATGTGTGACACATTTTCAAGACCAAGGATTTGCTCCATTAATTCGACAG 493
 QY 831 -----ACGTGTAAACAGGTGTGCTCCGACGAAACGCTGTGACGCTGCGCAACTTC 883
 Db 494 GATTTCTGCTGTAAATATGCTGCTGCGACAGCTCCAGAGCTTCTGCTTCCAC 553
 QY 884 GTTACTCTCAAGAAAGATTCGCAACCGAGATATCATTTGATGTACAGACTGATATCA 943
 Db 554 GATTGATCAAGAGCGATCAATTAATTCGATTAATGCGTGGACGTCAGTACGCTTTTA 613
 QY 944 ACACCCCTTGAATTAAGCGATGAATCAAGCTTTGCCAGTTGAT 987
 Db 614 GCACGGTTGATATACCGAGAGAAATCCGAACTTCTCAGCAAAAT 657

RESULT 13
 AAH24833/C
 ID AAH24833 standard; DNA; 813 BP.

AAH24833;

22-AUG-2001 (first entry)

Nucleotide sequence of antisense strand of cuticlin open reading frame.

Cuticlin; gene therapy; vaccine; helminth parasite; ss.

Dirofilaria immitis.

US6248329-B1.

19-JUN-2001.

01-JUN-1999; 99US-0323427.

01-JUN-1998; 98US-0087435.

(CHAN/) CHANDRASHEKAR R.

(MORA/) MORALES T H.

Chandrashekar R, Morales TH;

WPI: 2001-396953/42.

Novel nucleic acid molecule encoding a Dirofilaria immitis cuticlin protein is useful as a vaccine to prevent parasitic helminth infection

Example 1; Column 49-50; 29pp; English.

The specification describes a Dirofilaria immitis nucleic acid molecule, that encodes a cuticlin protein. The Dirofilaria immitis nucleic acid molecule is useful as a probe to identify nucleic acid molecules, as a primer to produce nucleic acids, as a therapeutic reagent, e.g., gene therapy to inhibit cuticlin activity or production, or in a vaccine to prevent infection with helminth parasites. The cuticlin protein, antibodies raised against it, and inhibitory compounds of cuticlin may all be used in compositions to protect animals, especially mammals such as cats, dogs, and humans. The antibodies may be used to passively immunize an animal, or as reagents in assay to detect infection of helminths, or as tools to screen expression libraries to recover desired

CC proteins. They may also be used to target cytotoxic agents to the parasite and kill it directly. The present sequence is the complement of an open reading frame encoding a cuticlin polypeptide.

Sequence 813 BP; 254 A; 164 C; 148 G; 247 T; 0 other;

Query Match 26.7%; Score 310.4; DB 22; Length 813;
 Best Local Similarity 70.7%; Pred. No. 5.9e-77;
 Matches 455; Conservative 0; Mismatches 161; Indels 28; Gaps 2;

QY 372 GGAAGCTGATTAACAGTTAGTGC-ACAGATGAGTATCTGAATACAACTGCTTTG 430
 Db 800 GAAAGCTGATTAAGCCGTTAGTACGTGCTTGAATATCTGAATACAACTGATTC 741
 QY 431 AACTCAAAATTTGCCGATGAGTATGCGCTTATGAATTTTGGATGGACCAACCG 490
 Db 740 AAATCAATGTTGATCAATGCGCTTATGATGATGATGATGATGATGATGATGAT 681
 QY 491 GTCAACCAATTTGCTATCATTTGTCAGCCAGCTTATGATTAATGAGATGCGATT 550
 Db 680 GTCAACCTGTTGATTTGCAATGATGATGATGATGATGATGATGATGATGATGAT 621
 QY 551 CTGAACCTGATATCTTCTGCGGCTGTCATTCCTGCTTTGCTGATGATGATGAG 610
 Db 620 CAGAGACTACAGATCATTTCTGCTGATGATGATGATGATGATGATGATGATGATG 561
 QY 611 GTGATACTGAGAAATTTCAATGCTGATGATGATGATGATGATGATGATGATGATG 670
 Db 560 GTGATGCACTGAGATTTGATGATGATGATGATGATGATGATGATGATGATGATG 501
 QY 671 ATTTGGAATATCCAGATTTTAAATGCTGCGCAAGAAAGCTCAGTATACAAATATGCG 730
 Db 500 ATTTGGAATATTTATACAGATTTTAAATGCTGCGCAAGAAAGCTCAGTATTAATATG 441
 QY 731 ATGATACAGAGCTTTTATCAATGCGAGATGATGATGATGATGATGATGATGATGAT 790
 Db 440 ATGATACAGAGCTTTTATCAATGCGAGATGATGATGATGATGATGATGATGATGAT 381
 QY 791 AATGTGTGACACATTTTCAAGACCAAGAGATTCG----- 830
 Db 380 AATGTGTGACACATTTTCAAGACCAAGAGATTTTCCATTAATATGAGCAACA 321
 QY 831 -----ACGTGTAAACAGGTGTGCTCCGACGAAACCTGCTCAGCTGCTTCTCCAC 883
 Db 320 GATTTGCTGCTGTAATAATCTGCTGCTGACACGCTCCAGAAAGCTTCTTCTCCAC 261
 QY 884 GTTACTCTCAAGAAAGATCTGCAACCGGAAATATCATTTGATGTAGAACTGATATCA 943
 Db 260 GATTGATCAAGAGCGATCAATTAATCTGATTAATGCGTGGACGTACGTACGGTTT 201
 QY 944 ACACCCCTTGAATTAAGCGATGAATCAAGCTTTGCCAGTTGAT 987
 Db 200 GCACGGTTGATATACCGAGAGAAATCCGAACTTCTCAGCAAAAT 157

RESULT 14
 AAV21209/C
 ID AAV21209 standard; DNA; 1664976 BP.

AAV21209;

10-NOV-1998 (first entry)

Methanococcus jannaschii circular chromosome.

Methanococcus jannaschii; methanogenic archaeon; circular chromosome; genome; autotrophic; extrachromosomal element; identification; ds.

Methanococcus jannaschii.

W09807830-A2.

26-FEB-1998.

KM Ureaplasma urealyticum; nucleic acid detection; infection; pathogen;
 KM human urogenital tract; pregnancy; neonatal disease; drug therapy;
 KM suppurative arthritis; ss.
 XX
 OS Ureaplasma urealyticum.
 XX
 PN W0939007-A1.
 XX
 PD 05-AUG-1999.
 XX
 PF 29-JAN-1999; 99WO-US01972.
 XX
 PR 30-JAN-1998; 98US-0073189.
 XX
 PA (UABR-) UAB RES FOUND.
 XX
 PI Cassell GH, Chen EX, Glass JT, Glass JS, Helner CR;
 PI Letkowitz E;
 XX
 DR WPI: 1999-469343/39.
 XX
 PT Detection of Ureaplasma urealyticum using novel genes, probes and
 PT primers
 XX
 PS Claim 1; Page 41-46; 110pp; English.
 XX
 CC The present invention provides methods for the detection and diagnosis
 CC of Ureaplasma urealyticum infection. It provides novel genes
 CC (AA99501-681) that can be used as a source of primers and probes for the
 CC detection and/or quantification of U. urealyticum in a biological
 CC sample. The probes that can be used in the method of the invention by
 CC forming target:probe complex is complementary to a region selected from
 CC one of the 181 nucleic acid sequences (AA99501-681). U. urealyticum is
 CC an opportunistic pathogen of the human urogenital tract that is a
 CC significant cause of adverse pregnancy outcome, neonatal disease, and
 CC suppurative arthritis. As the infections are commonly asymptomatic, it is
 CC important to have specific and sensitive methods for detecting their
 CC presence in a patient. Also, as the pathogen has no current antibiotic
 CC directed specifically against it, it would be advantageous to isolate and
 CC detect gene sequences which are unique to it, and utilise these as a
 CC basis for diagnosis of U. urealyticum infection as well as to develop new
 CC and improved drug therapies. The present invention provides such novel
 CC polynucleotide sequences (AA99501-681).
 XX
 SQ Sequence 14066 BP; 4508 A; 1702 C; 1539 G; 6317 T; 0 other;
 QY
 Query Match 3.3%; Score 38.8; DB 20; Length 14066;
 Best Local Similarity 57.4%; Pred. No. 4.8;
 Matches 70; Conservative 0; Mismatches 52; Indels 0; Gaps 0;
 QY 81 TGAGCCAGAAATGTAATGTGCACTTCAATCAATCAATTTTAATACAGTAATGC 140
 DB 9196 TAAACCAAAATGTAATGTGCACTTCAATCAATCAATTTTAATACAGTAATGC 9137
 QY 141 ATTGGAAGCATGTTTATGTGTAAGGCTTTATGATCAAGAAGTTCGCTAATGATGA 200
 DB 9136 TATTAAAGAGTAGTGCACATATATGCTCAAGTACATTTTAAGCTTAATAAATAATATGA 9077
 QY 201 AG 202
 DB 9076 AG 9075
 RESULT 17
 ID ABR39779/C
 XX ABR39779 standard; CDNA; 6799 BP.
 AC ABR39779;
 XX
 DT 21-MAY-2002 (first entry)
 XX
 DE CDNA encoding full length L1428P.
 XX

KM Lung tumour; cancer; T cell; immune response stimulator;
 KM cytosolic; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN W0200204514-A2.
 XX
 PD 17-JAN-2002.
 XX
 PF 10-JUL-2001; 2001WO-US22058.
 XX
 PF 11-JUL-2000; 2000US-0614124.
 PR 29-AUG-2000; 2000US-0651363.
 PR 08-SEP-2000; 2000US-0658824.
 PR 26-SEP-2000; 2000US-0671325.
 PR 06-OCT-2000; 2000US-0677419.
 PR 30-OCT-2000; 2000US-0702705.
 PR 13-DEC-2000; 2000US-0736457.
 PR 03-MAY-2001; 2001US-0849626.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Wang T, Matanabe Y, Henderson RA, Johnson JC, Retter KM;
 PI Marnerakis M, Carter D, Fanger GR, Vedrick TS, Bangur CS;
 PI McNabb A, Wang A, Fanger N, Switzer A, McNeill PD, Clapper JD;
 XX
 DR WPI: 2002-164634/21.
 DR P-PSDB; AAU85654, AAU85655, AAU85656, AAU85657, AAU85658, AAU85659,
 DR AAU85660.
 XX
 PT Novel polynucleotide encoding a lung tumour polypeptide useful for
 PT stimulating and/or expanding T cells specific for a tumour protein
 XX
 PS Claim 1; SEQ ID No 1883; 223pp; English.
 XX
 CC The invention describes an isolated polynucleotide and polypeptide
 CC useful for stimulating and/or expanding T cells specific for a tumour
 CC protein for determining the presence of a cancer in a patient. A
 CC composition containing the polynucleotide and/or polypeptide is useful
 CC for treating a lung cancer in a patient. The polypeptide is useful for
 CC removing tumour cells from a biological sample. The polynucleotide is
 CC also useful as probe or primer to detect the level of mRNA encoding a
 CC tumour protein. This sequence encodes a lung tumour associated protein
 CC or protein fragment, described in the method of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 6799 BP; 1801 A; 1865 C; 1539 G; 1594 T; 0 other;
 QY
 Query Match 3.3%; Score 38; DB 24; Length 6799;
 Best Local Similarity 59.1%; Pred. No. 5.8;
 Matches 65; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
 QY 235 TTGTGATTCGAAAGTGGCGGTACAGCATCTGAAATCAACGAGTATTTTGTACAA 294
 DB 5002 TTTGTACAAAGCAAAATGAGATGCCAGCATCTCAATCGCAGATGTTCTCTGCA 4943
 QY 295 ACAAGCTGTGATTTTGGTTTCATCATATTTTACCAAGTTGATCG 344
 DB 4942 GTACCTGTGCTGATTTTCTTCTAGATGATTTATCTTTATTTATGG 4893
 RESULT 18
 ID ABO70961/C
 XX ABO70961 standard; DNA; 6242 BP.
 AC ABO70961;
 XX
 DT 29-AUG-2002 (first entry)
 XX
 DE Listeria monocytogenes 4b contig DNA sequence #903.
 XX

XX Antibacterial; Listeria; food contamination; mutational analysis;
KM infection; ds.
XX
OS Listeria monocytogenes 4b.
XX
PN WO200228891-A2.
XX
PD 11-APR-2002.
XX
PF 04-OCT-2001; 2001WO-FR03061.
XX
PR 04-OCT-2000; 2000FR-0012697.
XX
PA (INSP) INST PASTEUR.
PA (CNRS) CNRS CENT NAT RECH SCI.
PI Kunst F, Glaser P;
PI
DR WPI; 2002-332479/37.
XX
PT New genomic sequences from Listeria species, useful for detection,
PT treatment and prevention of infection, also related polypeptides,
PT antibodies and modulators -
XX
PS Claim 14; SEQ ID 3774; 180pp; French.
XX
CC The present invention relates to nucleic acid sequences
CC (AB067188-AB071212) from Listeria sp. The sequences are useful as probes
CC and primers for identification and/or detection of Listeria (e.g. as
CC contaminants in foods, or mutational analysis) and for analysis of
CC gene expression. Proteins encoded by the nucleic acid sequences can be
CC used to screen for compounds that modulate gene expression, replication
CC and pathogenicity of Listeria (potential therapeutic agents), also for
CC treating infections by Listeria, and are useful as immunogens in
CC anti-Listeria vaccines.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 6242 BP; 1954 A; 1060 C; 1164 G; 2064 T; 0 other;
Query Match 3.2%; Score 37.2; DB 24; Length 6242;
Best Local Similarity 53.4%; Pred. No. 9.3;
Matches 78; Conservative 0; Mismatches 68; Indels 0; Gaps 0;
QY 821 AAGGATTCGAGCTGTTAAACAGGTGGTCCCGACCAAACTGCTGACGCTGGCGAAC 880
DB 3172 AAGAAAGGTGGAACGGTTACAGCCGCAATCTTCGGTATCAATGATGAGCTCCGCAA 3113
QY 881 TTGGTTTACACAGAAAGATCTGCAGAACCGGAGATATCATTTGATGTCAGAACTGATA 940
DB 3112 TTATTTTAAAGTCGAAAGAAAGAGCGAGTCGGGAAACATTCCTATATCGCAACATCA 3053
QY 941 TCAACACCCCTGAAATTAAGCATGAT 966
DB 3052 AAGTAACCTCTGAAGTTGGTGTGAT 3027
RESULT 19
AA216979/c
ID AA216979 standard; cDNA; 783 BP.
XX
AC AA216979;
XX
DT 12-OCT-1999 (first entry)
XX
DE Human gene expression product cDNA sequence SEQ ID NO:449.
XX
KM Human; gene: gene expression product; diagnosis; therapy; probe;
KM detection; mapping; tissue typing; profiling; forensic; cancer;
KM genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX

OS Homo sapiens.
XX
PN WO9938972-A2.
XX
PD 05-AUG-1999.
XX
PF 28-JAN-1999; 99WO-US01619.
XX
PR 03-APR-1998; 98US-0080666.
PR 28-JAN-1998; 98US-0072910.
PR 24-FEB-1998; 98US-0075954.
PR 31-MAR-1998; 98US-0080114.
PR 03-APR-1998; 98US-0080515.
XX
PA (CHIR) CHIRON CORP.
PA (HYSE-) HYSEO INC.
XX
PI Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
PI Escobedo J, Garcia PD, Garcia V, Gliese K, Innis MA;
PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;
PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
PI Stache-Crain B, Sudduth-Klinger J, Williams LT;
XX
DR WPI; 1999-494092/41.
XX
PT Novel human genes and their expression products which are
PT differentially expressed in different cell types
XX
PS Claim 1; Page 2107; 2479pp; English.
XX
CC The present invention describes a library of human polynucleotides
CC comprising the sequences given in AA212532 to AA217779. Also described is
CC a method of detecting differentially expressed genes correlated with the
CC cancerous state of a mammalian cell, comprising detecting at least one
CC differentially expressed gene product in a test sample from a cell
CC suspected of being cancerous, where the gene product is encoded by one
CC of the 5248 polynucleotide sequences given in AA212532 to AA217779. The
CC polynucleotides can be used as a source of primers and probes, which can
CC be used for a variety of purpose, e.g. detection of expression levels,
CC mapping, tissue typing or profiling, forensics, genetic analysis and
CC detection of polymorphisms. Polypeptides encoded by the polynucleotides
CC can be used for raising antibodies for experimental, diagnostic and
CC therapeutic purposes. The polynucleotides may also be used to construct
CC arrays for diagnostics (which may be used to determine function of an
CC encoded protein), and to detect differences in expression levels between
CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
CC identify a genetic predisposition or susceptibility to a disease such as
CC cancer). The polynucleotides of the invention are especially used in the
CC diagnosis, prognosis and management of colorectal cancer, breast cancer,
CC and lung cancer. The polynucleotides can also be used to screen for
CC peptide analogues and antagonists.
XX
SQ Sequence 783 BP; 243 A; 176 C; 145 G; 197 T; 22 other;
Query Match 3.2%; Score 37; DB 20; Length 783;
Best Local Similarity 53.0%; Pred. No. 4.1;
Matches 79; Conservative 0; Mismatches 70; Indels 0; Gaps 0;
QY 588 CTGCTTTGCGATGATGTAACGGTGTACTGTGAAATCTTAATGCTGATGATGTGC 647
DB 392 CTGCTGAGAAAGAGCACTGATGTGAGTATGCATTAACCTCAATGCTGTTAACTGGC 333
QY 648 TCTTGATTAATATTGCTAATAATATTGGAATATCCACAGATTATATGCTGGCCAGA 707
DB 332 AGTTTCAATATGTGTATTAATTAACCTTGAATTAACCCATTGACGATGAGTGAAGG 273
QY 708 AGCTCAGTATACAAATATGCGATGAT 736
DB 272 GGGTCCCTTATATCATTTCTTGTAT 244
RESULT 20
AA299054/c

QY	588	CTCTTTTTCGATCATGGTACCGGTGATACGTGTGGAAATCTTAATCTCATGTGATTCG	64.7
Db	629	CTGCTGGAAAGAAAGAACTGATGTGTGGAGTATGCATTAATCTCATCTCTGTATTAACGCG	57.0
QY	648	TCTTGATTAATATTTGCTAAATATATTGGAAATATCCAAAGATTAAATGCTGTGGCCAGA	70.7
Db	569	AGTTTCACAAATATGTGATCTTTTACTTGATATTAACCCATTCGACTAAGGAGTATGACAAAG	51.0
QY	708	AGCTACGCTATACAAATATGGGATTCGAT	73.6
Db	509	GGGTCTCTGTATTCATTTGCTCTTGAT	48.1

Query Match	Best Local Similarity	Score	DB	Length
Matches	82; Conservative	0;	Mismatches	75; Indels
				Gaps
69	TGCTGTGCAAGTGAGCCGAGAAATGTAATGTGACCACTTCATTAATCAATTTTAA	128		
1030	TGCTGTGCAAGGAAAGACGCGAGTACGATTATAGACTTTTAATGGAAGCGGATTTTAA	971		
129	TACACGTAATGACATTCGAAAGACATGTTTATGTGAAAGCTTTATGATCAAGAAGCTTG	188		

Db 970 TGTTCGTGATATTTTATAGCAGATGGCCAGGCCATTATTTATGACAGACCAGAGATTT 9111
 189 CCGTAATGATGACGAGTGGACGTCAAGTTGGCCGAA TT 225
 Db 910 CCACAAAGTACACAGAGGTTGAAAGGCGGCTGGAA TT 874

RESULT 26
AA89354/c
ID AA89354 standard; cDNA: 6571 BP.
XX
AC AA89354;
XX
DT 23-APR-2001 (first entry)
XX
DE Arabidopsis thaliana silencing gene cDNA.
XX
KW Gene silencing; silencing gene; MOM; ss.
XX
OS Arabidopsis thaliana.
XX
FT Key Location/Qualifiers
FT CDS 310..635
FT /tag= a
XX
XX MO200100801-A2.
XX
PD 04-JAN-2001.
XX
PF 21-JUN-2000; 2000WO-EP05761.
XX
PR 23-JUN-1999; 99GB-0014623.
XX
PA (NOVS) NOVARTIS AG.
PA (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
PI Habu Y, Mittelsten Scheid O, Amedeo P, Paszkowski J;
XX
DR WP1: 2001-137952/14.
DR P-PSDB; AAB20062.
XX
PT Novel gene encoding a protein that controls gene silencing, in
PT particular silencing of plant genes -
XX
PS Claim 6; Page 22-32; 48pp; English.
XX
CC The present sequence is that of Arabidopsis thaliana ecotype Zurich
CC epigenetic silencing gene MOM cDNA clone. It encodes a 2001-amino
CC acid protein (see AAB20062) with similarity with AtPase/helicase
CC proteins of the SWI2/SNF2 family, which controls gene silencing.
CC particularly plant gene silencing. Trans-acting modified loci were
CC identified by T-DNA insertion mutagenesis using Arabidopsis thaliana
CC ecotype Zurich, which carries a heritable inactivated, methylated
CC hydrogymylin resistance gene (hgy). Mutation of a silencing
CC modified locus results in release of silencing of the hgy gene
CC and restores hygromycin resistance. Plants homozygous for the
CC silent resistance gene were subjected to transformation with a bar
CC selectable marker gene under control of the T-DNA 1'-2' dual
CC promoter. Transformants were selected and their progeny screened
CC for hygromycin resistance. The mutant phenotype (hygromycin
CC T-DNA insert). Cloning of the tagged gene allowed characterization
CC of the present sequence. 2 Allelic differences were found in the
CC corresponding gene of A. thaliana ecotype Columbia (see AA89385).
CC Gene silencing is useful as a molecular tool for regulating gene
CC expression.
XX
XX Sequence 6571 BP; 2060 A; 1347 C; 1470 G; 1694 T; 0 other;

Query Match 3.2%; Score 36.8; DB 22; Length 6571;
Best Local Similarity 48.1%; Pred. No. 12;
Matches 104; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

592	TTTGTCATGATGCTAACGCTGATCTGTGTAATTTCTAAATGCTGATGGATTCGCTCTT	651							
matches	11;	conservative	0;	Mismatches	124;	Indels	0;	Gaps	0;

Best Local Similarity	50.9%;	Pred. No. 16;			
Matches	87; Conservative	0; Mismatches	84; Indels	0; Gaps	0;

527 TTTATCAATAAGGACATGGGATTCGAAACCGTTGATACTTCTGCGCGGTTGTCATT 586


```

Db 1782 TTTTGTGTAATAGTATTTTATGTTTATTTTATTTTGGGTTAGTGCAGT 1841
      ||| ||||| | | ||| ||| ||| ||| ||| |||
QY 587 CCGGCTTTGCGATGATGTAACGGTACTGCGAAATTCATAAGCGATGATGTCG 646
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1842 ATTTGTTTTTATGATGATGTTACGTTAGGAATATTAATAATTTAAATTTAGTGTGTTT 1901
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 647 CTTGTGTAATATTTGCTTAATATTTGGAATATCCACAGATTTTAATGCG 697
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1902 TAATTTTATATTTGTTTACAGATTTGGTTGGCATATGATTTTATTTGCG 1952
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 29
ABQ67197
ID ABQ67197 standard; DNA; 1163020 BP.
XX
AC ABQ67197;
XX
DT 29-AUG-2002 (first entry)
XX
DE Listeria innocua contig DNA sequence #10.
XX
KW Antibacterial; Listeria; food contamination; mutational analysis;
XX infection; ds.
XX
OS Listeria innocua.
XX
PN WO200228891-A2.
XX
PD 11-APR-2002.
XX
PF 04-OCT-2001; 2001MO-FR03061.
XX
PR 04-OCT-2000; 2000FR-0012697.
XX
PA (INSP ) INST PASTEUR.
XX (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI Kunst F, Glaser P;
XX
DR WPI; 2002-332479/37.
XX
PT New genomic sequences from Listeria species, useful for detection,
XX treatment and prevention of infection, also related polypeptides,
XX
PT antibodies and modulators -
XX
PS Claim 5; SEQ ID 10; 180pp; French.
XX
XX
CC The present invention relates to nucleic acid sequences
CC (ABQ67188-ABQ71212) from Listeria sp. The sequences are useful as probes
CC and primers for identification and/or detection of Listeria (e.g. as
CC contaminants in foods, or mutational analysis) and for analysis of
CC gene expression. Proteins encoded by the nucleic acid sequences can be
CC used to screen for compounds that modulate gene expression, replication
CC and pathogenicity of Listeria (potential therapeutic agents), also for
CC treating infections by Listeria, and are useful as immunogens in
CC anti-Listeria vaccines.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX
SQ Sequence 1163020 BP; 388339 A; 197589 C; 235612 G; 341474 T; 6 other:

Query Match 3.1%; Score 36.4; DB 24; Length 1163020;
Best Local Similarity 49.5%; Pred. No. 1.6e+02;
Matches 94; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 36 TATTCATTTGCTTATTCGATTCGCGTTGACAAATGCTGCAAGGTGAGCCAGAAATTGA 95
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 311654 TGTTCGCTATATGTTTATATACAGTATTTTGGTGTGCAAGAAAAGCCAGATGA 311713
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 96 ATGTGACCAACTTCATAATCAATTTTAATACAGTAATGCAATTCGCAAGACATGT 155
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```

Db 311714 GCTATTGGAACTTTTAATGGAAGCGGATATTGATGTTGCGACATTTTACAGCAAGATGG 311773
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 156 TTATGTGAAGAGGCTTTATGATGCAGAAAGGTGCGCTAATGATGGAAGTGGAGTCAAGT 215
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 311774 CCAAGCGATTTATTTATGCAGAACCGAAGATTTCCACAAAGATACAGAAAGTTGAAAGC 311833
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 216 TGCCGCAATT 225
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 311834 GGCTGCAATT 311843
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 30
ABQ69245
ID ABQ69245 standard; DNA; 3011208 BP.
XX
AC ABQ69245;
XX
DT 29-AUG-2002 (first entry)
XX
DE Listeria innocua DNA sequence #684.
XX
KW Antibacterial; Listeria; food contamination; mutational analysis;
XX infection; ds.
XX
OS Listeria innocua.
XX
PN WO200228891-A2.
XX
PD 11-APR-2002.
XX
PF 04-OCT-2001; 2001MO-FR03061.
XX
PR 04-OCT-2000; 2000FR-0012697.
XX
PA (INSP ) INST PASTEUR.
XX (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI Kunst F, Glaser P;
XX
DR WPI; 2002-332479/37.
XX
PT New genomic sequences from Listeria species, useful for detection,
XX treatment and prevention of infection, also related polypeptides,
XX
PT antibodies and modulators -
XX
PS Claim 5; SEQ ID 2058; 180pp; French.
XX
XX
CC The present invention relates to nucleic acid sequences
CC (ABQ67188-ABQ71212) from Listeria sp. The sequences are useful as probes
CC and primers for identification and/or detection of Listeria (e.g. as
CC contaminants in foods, or mutational analysis) and for analysis of
CC gene expression. Proteins encoded by the nucleic acid sequences can be
CC used to screen for compounds that modulate gene expression, replication
CC and pathogenicity of Listeria (potential therapeutic agents), also for
CC treating infections by Listeria, and are useful as immunogens in
CC anti-Listeria vaccines.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX
SQ Sequence 3011208 BP; 941651 A; 568176 C; 559189 G; 942192 T; 0 other:

Query Match 3.1%; Score 36.4; DB 24; Length 3011208;
Best Local Similarity 49.5%; Pred. No. 2.2e+02;
Matches 94; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 36 TATTCATTTGCTTATTCGATTCGCGTTGACAAATGCTGCAAGGTGAGCCAGAAATTGA 95
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 401918 TGTTCGCTATATGTTTATATACAGTATTTTGGTGTGCAAGAAAAGCCAGATGA 401977
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 96 ATGTGACCAACTTCATAATCAATTTTAATACAGTAATGCAATTCGCAAGACATGT 155
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 401978 GCTATTGGAAGCTTTAATGGAAGGGAATATTGATGTTGCGACATTTTACAGCAAGATGG 402037
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

OY 156 TTATGTGAAAGCTCTTTATGATCAAGAGCTTGCCGTAATGATGAGTGCAGCTCACT 215
 DB 402038 CCAAGCGATTATTATGACAGACAGAGATTTCACCAAGATCAAGAGGTTTGAACG 402097
 OY 216 TGGCCGGAATT 225
 DB 402098 GCGTGGGAATT 402107

RESULT 31

ABA48451/C
 ID ABA48451 standard; DNA; 510 BP.

XX ABA48451;

XX 01-FEB-2002 (first entry)

XX Human breast cell single exon nucleic acid probe #7146.

KW Human; microarray; single exon probe; gene expression; breast;
 KW disease; cancer; ss.

OS Homo sapiens.

PN WO200157271-A2.

XX 09-AUG-2001.

PE 30-JAN-2001; 2001WO-US00662.

XX 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-496933/54.

Claim 4; SEQ ID NO 7146; 327pp + sequence listing; English.

CC The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human breast and BT 474 cells. The method involves contacting
 CC the probes with a collection of detectably labelled nucleic acids
 CC derived from mRNA of human breast, and then measuring the label
 CC bound to each probe of the microarray. The probes are useful for
 CC verifying the expression of regions of genomic DNA predicted to
 CC encode proteins. They are useful for gene discovery, and for
 CC determining predisposition and/or prognosing breast disease. Gene
 CC expression analysis is useful for assessing the toxicity of chemical
 CC agents on cells. The microarray of this invention presents a far greater
 CC diversity of probes for measuring gene expression, with far less bias
 CC than expressed sequence tag microarrays. The method is suitable for
 CC rapid production of functional information from genomic sequence. The
 CC present sequence is a single exon nucleic acid probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 510 BP; 149 A; 184 C; 68 G; 109 T; 0 other;

Query Match 3.1%; Score 36.2; DB 22; Length 510;

Best Local Similarity 46.3%; Pred. No. 5.7;
 Matches 119; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

OY 476 ATGGTGACCAACCGGTCAACAGTTGCTATCTATGCTGACGACGTTATACATA 535
 DB 414 ATGCTGATGACGAGCGTGTGATGCTGATGCTGATGCTGATGCTGATGCTGATG 355
 OY 536 AATGACATGCGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATG 595
 DB 354 ATGTTGAACGACGCTGCTGAAAAAGGTGTTGATGCTGATGCTGATGCTGATGCTGATG 295
 OY 596 TCGATGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATG 655
 DB 294 ATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATG 235
 OY 656 AATATTGCTAATATATTTGCAATATCCACAGTTTATGCTGCTGCCAAGAGCTCAG 715
 DB 234 CAGGTGCTGAAAAAGGTGTTGATGCAAGTGCAGATGCTGATGCTGATGCTGATGCTGATG 175
 OY 716 TATACAAATATGCGGAT 732
 DB 174 CTGTTAGTATGCTGCTGT 158

RESULT 32

ABA66355/C
 ID ABA66355 standard; DNA; 510 BP.

XX ABA66355;

XX 01-FEB-2002 (first entry)

XX Human foetal liver single exon nucleic acid probe #14660.

KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.

OS Homo sapiens.

PN WO200157277-A2.

XX 09-AUG-2001.

PE 30-JAN-2001; 2001WO-US00669.

XX 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483447/52.

Claim 4; SEQ ID NO 14660; 639pp + sequence listing; English.

CC The invention relates to a single exon nucleic acid probe for
 CC measuring human gene expression in a sample derived from human foetal
 CC liver. The single exon nucleic acid probes may be used for predicting,
 CC measuring and displaying gene expression in samples derived from human
 CC foetal liver. The present sequence is a single exon nucleic acid
 CC probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

from the human heart via microarrays. By measuring gene expression, the

PT single exon nuclear acid probes for analyzing gene expression in human brains -

```

xx Example 4; SEQ ID NO: 14763; 650pp + Sequence Listing; English.
xx
xx The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention.
xx
xx Sequence 510 BP; 149 A; 184 C; 68 G; 109 T; 0 other;
SQ
Query Match          3 1%; Score 36.2; DB 22; Length 510;
Best Local Similarity 46.3%; Pred.No.5.7;
Matches 119; Conservative 0; Mismatches 138; Indels 0; Gaps 0;
OY 476 ATGTGTCACCAACCAGCTTCAACCGATTCAATTTGGTATCATTTGGTGACGCCAGTTTAACATA 535
DB 414 ATGCTGATCAGCAGGCCTGTACTGTATGTCAGGATTTGGTGTGGTGACAGATGTTGAACG 355
OY 536 AATGACATGACGATTTCTGAAACCGTTGATCTTCTGCGCGGTGTCATTCCTGCTTG 595
DB 354 AGTTGTAAGACAGAGTGTCTGAAAAGGTTGATGACGAGTCACATGCTGATGCTGTAAGT 295
OY 596 TCGATGATGCTGAACGGTATCTACTGTGGAATTTCTAATGCTGATGAGTGGCTTTGATA 655
DB 294 ATGCTGATGCTGTGATGATGCTGTGTCAGATGTTGATGACAGATGTGAAGCAGGTGTAAG 235
OY 656 AATATTTGCTAAATATTGGAATATCCAACAGATTTAATGGCTGGCCAAAGCTCACG 715
DB 234 CAGGTGCTGAAAAAATGTTGATGCAAGTCAGATGCTGATCTGTAGTAGTGATG 175
OY 716 TATACAATATGCGGAT 732
DB 174 CTGTAAGTAGTCTGCT 158

```

RESULT 35
AAK40512/C
ID AAK40512 standard; DNA; 510 BP.
XX
XX AAK40512;
XX AC
XX 06-NOV-2001 (first entry)
XX DT
XX DE Human bone marrow expressed single exon probe SEQ ID NO: 15069.
XX XX
XX KW Human; bone marrow expressed exon; gene expression analysis; probe;
XX microarray; cancer; Leukaemia; Lymphoma; myeloma; ss.
XX OS
XX Homo sapiens.
XX PN WO200157276-A2.
XX PD
XX 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00668.
XX PR
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024253.
XX PR
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488900/53.
XX

PT	Human genome-derived single exon nucleic acid probes useful for
PT	analyzing gene expression in human bone marrow -
XX	
PS	Example 4; SEQ ID NO: 15069; 658pp + Sequence Listing; English.
XX	
CC	The present invention provides a number of single exon nucleic acid
CC	probes which are derived from genomic sequences expressed in the human
CC	bone marrow. They can be used to measure gene expression in bone marrow
CC	samples, which may enable the improved diagnosis and treatment of cancers
CC	such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC	the probes of the invention.
XX	
SO	Sequence 510 BP; 149 A; 184 C; 68 G; 109 T; 0 other;
Query Match	3.1%; Score 36.2; DB 22; Length 510;
Best Local Similarity	46.3%; Pred. No. 5.7;
Matches 119; Conservative	0; Mismatches 138; Indels 0; Gaps 0;
QY	476 ATGGTGAGCACCACCGGTCTACACAGTTTCATTTGCTATCTATTTGGTCAGCCAGTTTATCAT 535
DB	414 ATGCTGATGACAGGCGCTGTGATGTCAGGATTTGTTGGTGCAGATGTTGATGACG 355
QY	536 AATGAGCATGCGATTTCTGAAACCGTTGATCTTCTGCGCGGTTGCGCATTCCTGCTTG 595
DB	354 ATGTTGAGACAGAGTCTGTGAAAAGTGTGATGACAGGTGCAGATGCTGATGCTGATGTG 295
QY	596 TCGATGATGATGATGACGGTATCTGTAATTTCTAATGCTGATGATGATGCTGCTTGCATA 655
DB	294 ATGCTGATGATGATGATGATGCTGTGCGAGATGTTGATGACAGATGTTAAGCAGGTGTTGATG 235
QY	655 AATATTTGCTAATATATTTTGGATATCCACAGATTTAATGCTGCGCCAGAGCTCAGC 715
DB	234 CAGGCGCTGAAAATGTTGTTGATGCAAGTGCAGATGCTGATGCTGTGATGCTGATG 175
QY	716 TATACAAATATTCGGAT 732
DB	174 CTGTTAGTGTGCTGCT 158
RESULT 36	
AA146546/C	
ID	AA146546 standard; DNA; 510 BP.
XX	
AC	AA146546;
XX	
DT	17-OCT-2001 (first entry)
XX	
DE	Probe #1532 used to measure gene expression in human placenta sample.
XX	
KW	Probe; microarray; human; placenta; antenatal diagnosis;
KW	genetic disorder; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200157272-A2.
XX	
PD	09-AUG-2001.
XX	
PE	30-JAN-2001; 2001MO-US000663.
XX	
PR	04-FEB-2000; 2000US-0180312.
PR	26-MAY-2000; 2000US-0207456.
PR	30-JUN-2000; 2000US-0608408.
PR	03-AUG-2000; 2000US-0632366.
PR	21-SEP-2000; 2000US-0234687.
PR	27-SEP-2000; 2000US-0236359.
PR	04-OCT-2000; 2000GB-0024263.
XX	
PA	(MOLE-) MOLECULAR DYNAMICS INC.
PI	Penn SG, Hanzel DK, Chen W, Rank DR;
XX	
XX	WPI; 2001-48897/53.

Query Match	3.1%	Score 36.2	DB 22	Length 510
Best Local Similarity	46.38	Pred. NO. 5.7	138	Indels 0
Matches 119	Conservative 0	Mismatches 138	Indels 0	Gaps
QY 476	ATGCTGGACCAACCGGTCAACCAAGTTCATCTATCATTTGGTCAGCCAGATTATCATATA	535		
Db 414	ATGCTGATGACGAGCGGCTGATGCTGATGACAGATATTTGGTGGTGGCAGATGTGATGACAG	355		
QY 536	AATGGACATGGCATTTCTGAACCGTCTGATATCTTTCTCGCGGCTGTCCATTCCTGCTTG	595		
Db 354	ATGTTGAAGCAGGCTGCAAAAAGTGTGATGACAGGTGACAGTGTGATGCTGTGATG	295		
QY 596	TCGATGATGTTACGATACGTCATCTGTGGAATTTCTAAATTCGTATGATGATGCTGCTTGATA	655		
Db 294	ATGCTGATGATGATGATGATGATGCTGTGTCACATGTTGATGACAGATGTTGAAGCAGTGTGATG	235		
QY 656	AATATTGCTTAATTAATTTGGGAATATCCACAGATTTTAATGCTGCGCCAGAGAGCTCAG	715		
Db 234	CAGGTGCTGGAATAATGTTGTTGATGCAAGTGCATGCTGATGCTGATGATGATGCTGATG	175		
QY 716	TATCAAAATATGCGGAT 732			
Db 174	CTGCTAGTATGATGCTGCT 158			
RESULT 37				
ID ABK77860	standard; DNA; 1103 BP.			
XX ABK77860;				
XX 13-AUG-2002 (first entry)				
DE Bacillus clausii genomic sequence tag (GSR) #703.				
XX Differential gene expression: genomic sequenced tag; GSR;				
KW altered culture condition: environmental stress;				
KM physiological provocation; ds.				
XX Bacillus clausii.				
OS WO200229113-A2.				
PN 11-APR-2002.				
XX 05-OCT-2001; 2001WO-US31437.				
PF 06-OCT-2000; 2000US-0680598.				
PR 27-MAR-2001; 2001US-279526P.				
XX (NOVO) NOVOZYMES BIOTECH INC.				
PA (NOVO) NOVOZYMES AS.				
PI Berka R, Clausen IG;				
XX WPI: 2002-416684/44.				
XX Monitoring differential expression of several genes in first Bacillus				
PT cell relative to expression of same genes in one or more second				
XX Human genome-derived single exon nucleic acid probes useful for				
PT analyzing gene expression in human placenta -				
XX Claim 25; SEQ ID NO 15232; 654bp; English.				
XX The present invention relates to single exon nucleic acid probes (SENP)				
CC The present sequence is one such probe. The probes are useful for				
CC producing a microarray for predicting, measuring and displaying gene				
CC expression in samples derived from human placenta. The probes are useful				
CC for antenatal diagnosis of human genetic disorders.				
XX Sequence 510 BP; 149 A; 184 C; 68 G; 109 T; 0 other;				

PT	Bacillus cells, by using substrate containing Bacillus genomic
PN	sequenced tag array -
PS	
XX	Claim 11; SEQ ID NO 5151; 200pp; English.
CC	The invention describes a method of monitoring differential expression of
CC	genes in a first Bacillus cell relative to expression of the genes in
CC	other Bacillus cells, comprising hybridising labelled nucleic acid probes
CC	isolated from Bacillus cells to a substrate containing array of Bacillus
CC	genomic sequenced tags (GST), examining the array, and determining
CC	relative gene expression by an observed hybridisation reporter signal of
CC	a spot in the array. The method is useful for measuring the expression of
CC	genes in a first Bacillus cell relative to expression of the same genes
CC	in one or more second Bacillus cells. The method is useful for monitoring
CC	global expression of several genes from a Bacillus cell, discovering new
CC	genes, identifying possible functions of unknown open reading frames and
CC	monitoring gene copy number variation and stability. Monitoring changes
CC	in expression of genes may be used to provide a representation of the way
CC	in which Bacillus cells adapt to changes in culture conditions,
CC	environmental stress or other physiological provocation. Extensive
CC	follow-up characterisation is unnecessary, when one spot on an array
CC	equals one gene or one open reading frame, since sequence information is
CC	available. This sequence represents a genomic sequence tag (GST) used in
CC	the method of the invention.
CC	Note: The sequence data for this patent did not form part of the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at
CC	ftp.wipo.int/pub/published_pct_sequences.
XX	
SQ	Sequence 1103 BP; 232 A; 272 C; 270 G; 329 T; 0 other;
	Query Match 3.1%; Score 36.2; DB 24; Length 1103;
	Best Local Similarity 56.2%; Pred. No. 8.1;
	Matches 68; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
OY	471 TTGTGATGCTGGACCAACCGGTCACCAAGTTGAATTTGCTATCATTTGGTCAGCAATTTA 530
DB	812 TTTTCAGGATGGTGGTTCTTACGTCATTTTCCTTATTTTGTACCTTAAATGGGCAATCGG 871
OY	531 TCATTAATGAGCATCGATTCGTAACCGCTTATATCTTCTGCGGGTGTGCATTCCTG 590
DB	872 TCATCGCGGATTTTGCATTTTGATTCCTCTGTCTTTTCCCGCTGTATCCGACCTG 931
OY	591 C 591
DB	932 C 932
RESULT 38	
ASAS45434	
ID	ASAS45434 standard; DNA: 7810 BP.
XX	
XX	ASAS45434;
DT	18-DEC-2001 (first entry)
DE	Chemically pretreated genomic DNA associated with cell cycle #70.
XX	
RW	Cell cycle; human; Cpg dinucleotide; cytosine methylation; HIV; aging;
RW	human immunodeficiency virus; neurodegenerative disorder; solid tumour;
RW	grft-versus-host disease; glomerular disease; Lewy body disease; cancer;
RW	arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritic;
RW	immunosuppressive; antitumour; cytosolic; antiarteriosclerotic; ds;
XX	PCR primer.
OS	
XX	Homo sapiens.
PN	WO200168911-A2.
XX	
PD	20-SEP-2001.
XX	
PF	15-MAR-2001; 2001WO-EP02945.
XX	

PR 15-MAR-2000; 2000DE-1013847.
 PR 06-APR-2000; 2000DE-1019058.
 PR 07-APR-2000; 2000DE-1019173.
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 PA (EPIG-) EPIGENOMICS AG.
 PI Olek A, Piepenbrock C, Berlin K;
 DR WPI: 2001-602751/68.
 XX
 PT Designing primers and probes for analysing diseases associated with
 PT cytosine methylation state e.g. arthritis, cancer, aging,
 PT arteriosclerosis comprising fragments of chemically modified genes
 PT associated with cell cycle -
 XX
 PS Claim 1: SEQ ID No 139; 28pp; English.
 XX
 CC Sequences AAS45296-AAS45520 represent chemically pretreated genomic DNA
 CC molecules associated with the cell cycle and specific PCR primers of the
 CC invention. The sequences are useful for detecting the methylation state
 CC of all CpG dinucleotides in a sequence and therefore for analysing
 CC associated diseases. By analysing cytosine methylations in the pretreated
 CC DNA, genetic and/or epigenetic parameters for the diagnosis and therapy
 CC of existing diseases or the predisposition to specific diseases can be
 CC ascertained. The parameters may be compared to another set of genetic
 CC and/or epigenetic parameters, the differences serving as basis for
 CC diagnosis and/or prognosis events which are disadvantageous to patients.
 CC The sequences of the invention are useful for the diagnosis and therapy
 CC of HIV infection, neurodegenerative disorders, graft-versus-host disease,
 CC aging, glomerular disease, Lewy body disease, arthritis,
 CC arteriosclerosis, solid tumours and cancers.
 CC
 SQ Sequence 7810 BP; 1928 A; 314 C; 2058 G; 3510 T; 0 other:
 XX
 Query Match 3.1%; Score 36.2; DB 22; Length 7810;
 Best Local Similarity 59.0%; Pred. No. 20;
 Matches 62; Conservative 0; Mismatches 43; Indels 0; Gaps 0;
 OY 235 TTGATTCATGCAATGTCGCGTACACGATCTCGAATCCACGNGATTTTGTATACA 294
 DB 6550 TATGTTTGGTTGATGATTCGTCGATTTAATGTTTGGAGCTGTGCGAATGTTGTATT 6609
 OY 235 ACACTGTTGTCATTTGCTTCATCCATTATTTGTACCAAGTT 339
 DB 6610 TAGACGTTTGATGATATTATTTATAGATTATTTGATAAAGATT 6654
 XX
 RESULT 39
 ABR28281
 ID ABR28281 standard; DNA: 7810 BP.
 XX
 AC ABR28281;
 XX
 DT 23-APR-2002 (first entry)
 XX
 DE DNA transcription associated genomic DNA #78.
 XX
 KW DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;
 KW PNA: cytosine methylation state; SNP; retroviral infection; gene; ds;
 KW single nucleotide polymorphism; adenosine deaminase deficiency; cancer;
 KW viral infection; Sezary syndrome; haematological disorder; tuberculosis;
 KW immunological disorder; Werner syndrome; developmental disorder;
 KW psoriasis; Rieger's syndrome; neurological disorder; erythropoiesis;
 KW neurodegenerative disorder; Mecklenburg syndrome; Niemann-Pick disease;
 KW myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;
 KW angiogenesis; congenital heart disease; HDR syndrome; gene therapy;
 KW polyglutamine disorder; solid tumour.
 XX
 OS Unidentified.
 XX
 PN WO200192565-A2.

XX
 PD 06-DEC-2001.
 XX
 XX 06-APR-2001; 2001WO-EP03973.
 XX
 PF
 XX
 PR 06-APR-2000; 2000DE-1019058.
 PR 07-APR-2000; 2000DE-1019173.
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 PA (EPIG-) EPIGENOMICS AG.
 PI Olek A, Piepenbrock C, Berlin K;
 DR WPI: 2002-090046/12.
 XX
 PT New nucleic acids or oligomers, useful for diagnosing or treating
 PT diseases associated with DNA transcription, e.g. immunological
 PT disorders, Werner syndrome, psoriasis, myocardial infarction, solid
 PT tumours or cancer -
 XX
 PS Claim 1: SEQ ID No 155; 32pp; English.
 XX
 CC The invention relates to a nucleic acid, which comprises a segment of the
 CC chemically pretreated DNA of genes associated with DNA transcription from
 CC one of 346 sequences, and an oligomer, in particular an oligonucleotide
 CC or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical
 CC to the chemically pretreated DNA of genes associated with DNA
 CC transcription. The set of oligomer probes are useful for detecting the
 CC cytosine methylation state and/or single nucleotide polymorphisms (SNPs)
 CC in a chemically pretreated genomic DNA. The nucleic acids are useful for
 CC diagnosing or treating diseases associated with DNA transcription
 CC (particularly with the methylation status), e.g. adenosine deaminase
 CC deficiency, viral infection, retroviral infection, Werner syndrome,
 CC haematological disorders, immunological disorders, Mecklenburg syndrome,
 CC tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,
 CC neurological disorders, neurodegenerative disorders, Mecklenburg
 CC syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial
 CC infarction, hypertension, angiogenesis, erythropoiesis, congenital heart
 CC disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours
 CC or cancer. Sequences ABR28127-ABR28472 represent DNA transcription
 CC associated genomic DNA molecules of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification but was obtained in electronic format directly from the
 CC European Patent Office.
 CC
 SQ Sequence 7810 BP; 1928 A; 314 C; 2058 G; 3510 T; 0 other:
 XX
 Query Match 3.1%; Score 36.2; DB 24; Length 7810;
 Best Local Similarity 59.0%; Pred. No. 20;
 Matches 62; Conservative 0; Mismatches 43; Indels 0; Gaps 0;
 OY 235 TTGATTCATGCAATGTCGCGTACACGATCTCGAATCCACGNGATTTTGTATACA 294
 DB 6550 TATGTTTGGTTGATGATTCGTCGATTTAATGTTTGGAGCTGTGCGAATGTTGTATT 6609
 OY 235 ACACTGTTGTCATTTGCTTCATCCATTATTTGTACCAAGTT 339
 DB 6610 TAGACGTTTGATGATATTATTTATAGATTATTTGATAAAGATT 6654
 XX
 RESULT 40
 AAH34068/C
 ID AAH34068 standard; cDNA: 1481 BP.
 XX
 AC AAH34068;
 XX
 DT 03-SEP-2001 (first entry)
 XX
 DE Human colon cancer antigen encoding cDNA SEQ ID NO:1150.
 XX
 KW Human: colon cancer; colon cancer antigen; diagnosis; detection;
 KW colorectal carcinoma; ss.

XX Homo sapiens.
XX
XX WO200122920-A2.
XX
XX 05-APR-2001.
XX
XX 28-SEP-2000; 2000WO-US26524.
XX
XX 29-SEP-1999; 99US-0157137.
PR 03-NOV-1999; 99US-0163280.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
DR WPI: 2001-235357/24.
XX P-PSDB; AAG74663.
XX
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX useful for preventing, diagnosing and/or treating colorectal cancers -
XX
XX Claim 1; Page 2962-2963; 9803pp; English.

AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon cancer-associated nucleic acid molecules (N) and proteins (P), where the proteins are collectively known as colon cancer antigens. The colon cancer antigens have cytostatic activity and can be used in gene therapy and vaccine production. N and P may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate P expression. For example, N and P may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of P by expressing inactive proteins or to supplement the patients own production of P. Additionally, N may be used to produce the colon cancer-associated PS, by inserting the nucleic acids into a host cell and culturing the cell to express the proteins. N and P can be used in the prevention, diagnosis and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204 and AAG77789 represent sequences used in the exemplification of the present invention.

N.B. Pages 666 to 662 and page 7053 of the sequence listing were missing at time of publication, meaning no sequences are present for SEQ ID NO:1027 to 1052, 7921 and 7922.

Sequence 1481 BP; 412 A; 259 C; 346 G; 459 T; 5 other;

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2003, 15:29:13 ; Search time 72 Seconds
(without alignments)
4945.163 Million cell updates/sec

Title: US-10-054-562a-3

Perfect score: 1161

Sequence: 1 atgatgattcgtctattgtc.....ttcgtccaatcgaagca 1161

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapept 1.0

Searched: 441362 seqs, 1533381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_NA:*
1: /cgn2_6/prodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/prodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/prodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/prodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/prodata/1/ina/PCUTUS.COMB.seq:*
6: /cgn2_6/prodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1161	100.0	1161	4	US-09-323-427-3
2	1161	100.0	1161	4	US-09-323-427-5
3	1161	100.0	1161	4	US-09-812-642-3
4	1161	100.0	1161	4	US-09-812-642-5
5	1161	100.0	1161	4	US-09-323-427-1
6	1161	100.0	1161	4	US-09-323-427-2
7	1161	100.0	1161	4	US-09-812-642-1
8	1161	100.0	1161	4	US-09-812-642-2
9	504.2	43.4	1372	4	US-09-323-427-6
10	504.2	43.4	1372	4	US-09-323-427-7
11	504.2	43.4	1372	4	US-09-812-642-6
12	504.2	43.4	1372	4	US-09-812-642-7
13	461	39.7	892	4	US-09-323-427-16
14	461	39.7	892	4	US-09-323-427-18
15	461	39.7	892	4	US-09-812-642-16
16	461	39.7	892	4	US-09-812-642-18
17	310.4	26.7	813	4	US-09-323-427-8
18	310.4	26.7	813	4	US-09-323-427-10
19	310.4	26.7	813	4	US-09-812-642-8
20	310.4	26.7	813	4	US-09-812-642-10
21	35.4	3.0	348	4	US-09-308-003-45
22	35.4	3.0	1182	4	US-09-308-003-8
23	34.2	2.9	4865	3	US-08-894-017-24
24	34	2.9	34	4	US-09-323-427-11
25	34	2.9	34	4	US-09-323-427-12
26	34	2.9	34	4	US-09-812-642-11
27	34	2.9	34	4	US-09-812-642-12

28	34	2.9	895	4	US-09-602-877A-91	Sequence 91, App1
29	34	2.9	1521	4	US-09-134-001C-1445	Sequence 1445, App
30	33.6	2.9	2427	4	US-09-134-001C-268	Sequence 268, App
31	33.4	2.9	1867	1	US-07-772-087-1	Sequence 1, App1
32	33.4	2.9	16950	4	US-09-453-702B-166	Sequence 166, App
33	33	2.8	11225	6	5182210-9	Patent No. 5182210
34	32.6	2.8	630	4	US-09-385-982-45	Sequence 45, App1
35	32.6	2.8	5177	6	5352450-1	Patent No. 5352450
36	32.4	2.8	576	5	PCT-US96-05320A-705	Sequence 705, App
37	32.4	2.8	4090	2	US-08-781-802-5	Sequence 5, App1
38	32.4	2.8	4090	4	US-08-694-078-5	Sequence 5, App1
39	32.4	2.8	31960	4	US-09-453-702B-11	Sequence 11, App1
40	32.2	2.8	1860	4	US-09-134-001C-1411	Sequence 1411, App
41	32.2	2.8	2121	4	US-09-134-001C-2071	Sequence 2071, App
42	32	2.8	1709	4	US-08-936-165A-188	Sequence 188, App
43	32	2.8	2338	1	US-08-425-069-1	Sequence 1, App1
44	32	2.8	2338	2	US-08-317-844B-1	Sequence 1, App1
45	31.8	2.7	12666	4	US-08-961-527-137	Sequence 137, App

ALIGNMENTS

```
RESULT 1
US-09-323-427-3
; Sequence 3, Application US/09323427
; Patent No. 6248329
; GENERAL INFORMATION:
; APPLICANT: Chandrasekhar, Ramaswamy
; APPLICANT: Morales, Tony H.
; TITLE OF INVENTION: Parasitic Helminth Cuticlin Proteins, Nucleic Acid
; FILE REFERENCE: HW-8
; CURRENT APPLICATION NUMBER: US/09/323.427
; EARLIER FILING DATE: 1999-06-01
; EARLIER APPLICATION NUMBER: 60/087,435
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1161
; TYPE: DNA
; ORGANISM: Dirofilaria immitis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1161)
US-09-323-427-3

Query Match      100.0%; Score 1161; DB 4; Length 1161;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGATGATTCGTCCTATTGCTTCTGACCTACTTATGCTATGCTTATTCGATTCGCG 60
    |||||||
Db 1 ATGATGATTCGTCCTATTGCTTCTGACCTACTTATGCTATGCTTATTCGATTCGCG 60

QY 61 GTTGACAAATGGTGTGCAAGGTGACCAAGAAATGGAATGACCAACTTCATTAATC 120
    |||||||
Db 1 GTTGACAAATGGTGTGCAAGGTGACCAAGAAATGGAATGACCAACTTCATTAATC 120

QY 121 AATTTTAAATACAGTAATGCTATGCAAGGACATTTATGGAAGGCTTTATGATCAA 180
    |||||||
Db 121 AATTTTAAATACAGTAATGCTATGCAAGGACATTTATGGAAGGCTTTATGATCAA 180

QY 181 GAAGTTGCGGTATGATGAAGGTGACGTCACGTCGCGGAATTCACCTTCATTTGAT 240
    |||||||
Db 181 GAAGTTGCGGTATGATGAAGGTGACGTCACGTCGCGGAATTCACCTTCATTTGAT 240

QY 241 TCATGCAATGTGTGGCGGTACACGATCTCTGAATCAGATGATTTTGTAAACAAC 300
    |||||||
Db 241 TCATGCAATGTGTGGCGGTACACGATCTCTGAATCAGATGATTTTGTAAACAAC 300

QY 301 GTTGTCATTTGCTTCAATCCATTAATTTGTTACCAAGTGAATGCTGATATCAGTACAA 360
    |||||||
```

```

Db 301 GTTGCATTCGTTTCATCCATTAATTTGTTACCAAGTTGATCGCATATCGATACAA 360
Qy 361 TGCCTTTACATGAGAGCTGATTAACACATTAAGTACAGATTTAGGATATCTGAATACCA 420
Db 361 TGCCTTTACATGAGAGCTGATTAACACATTAAGTACAGATTTAGGATATCTGAATACCA 420
Qy 421 ACTGCTTTCAACATCAATATGTCGCCGATGCGATGATGCGCTTATGAATTTGGATGCT 480
Db 421 ACTGCTTTCAACATCAATATGTCGCCGATGCGATGATGCGCTTATGAATTTGGATGCT 480
Qy 481 GGACCAACCGGTCAACCGATTTGCTATCATTTGGTCAAGCAGTTTATCATTAATG 540
Db 481 GGACCAACCGGTCAACCGATTTGCTATCATTTGGTCAAGCAGTTTATCATTAATG 540
Qy 541 ACATGCGATTTGAAACCGTGTGATCTTCTGCGCGGTGTCATTCCTGCTTTGTCGAT 600
Db 541 ACATGCGATTTGAAACCGTGTGATCTTCTGCGCGGTGTCATTCCTGCTTTGTCGAT 600
Qy 601 GATGCTAACCGGTGATCTGTTGGAATTTCTAATGCTGATGATGCTGCTTATGAATAT 660
Db 601 GATGCTAACCGGTGATCTGTTGGAATTTCTAATGCTGATGATGCTGCTTATGAATAT 660
Qy 661 TTGCTAATTAATTTGGAATTCACACAGATTTAATGCTGCGCAAGACTCACGTATAC 720
Db 661 TTGCTAATTAATTTGGAATTCACACAGATTTAATGCTGCGCAAGACTCACGTATAC 720
Qy 721 AAATATGCGGATGATCAACGCTTTCTATCAATGCGCAAGATTCATTAATTAAGAA 780
Db 721 AAATATGCGGATGATCAACGCTTTCTATCAATGCGCAAGATTCATTAATTAAGAA 780
Qy 781 CCAATATGCGAATGTTTTCACACAGATTTGAGAACCAAGATTCGAGCTGTTAAA 840
Db 781 CCAATATGCGAATGTTTTCACACAGATTTGAGAACCAAGATTCGAGCTGTTAAA 840
Qy 841 ACAGGTGTCGCCGACGAAACCTGCTGACGCTGCGCAACTGCTTACTCAAGAAAGA 900
Db 841 ACAGGTGTCGCCGACGAAACCTGCTGACGCTGCGCAACTGCTTACTCAAGAAAGA 900
Qy 901 TTGCGAAGACCGGAGATATCATTTGATGATGATGATGATGATGATGATGATGATGAT 960
Db 901 TTGCGAAGACCGGAGATATCATTTGATGATGATGATGATGATGATGATGATGATGAT 960
Qy 961 GATGATATCAACGCTTTGCGAGTTGATTTAGTACGCTGACGCTGCTGCTGCAACATTA 1020
Db 961 GATGATATCAACGCTTTGCGAGTTGATTTAGTACGCTGACGCTGCTGCTGCAACATTA 1020
Qy 1021 CAACGCTGATTAATCTGCTGACGATGAAATGAAATGATGATGATGATGATGATGATGAT 1080
Db 1021 CAACGCTGATTAATCTGCTGACGATGAAATGAAATGATGATGATGATGATGATGATGAT 1080
Qy 1081 ATGTTATGAGTTAAGCATTTGATGATGATGATGATGATGATGATGATGATGATGAT 1140
Db 1081 ATGTTATGAGTTAAGCATTTGATGATGATGATGATGATGATGATGATGATGATGAT 1140
Qy 1141 TTTTCGTCACAAATCAGAAAGCA 1161
Db 1141 TTTTCGTCACAAATCAGAAAGCA 1161

```

```

RESULT 2
US-09-323-427-5/c
; Sequence 5, Application US/09323427
; Patent No. 6248329
; GENERAL INFORMATION:
; APPLICANT: Chandrasekar, Ramaswamy
; APPLICANT: Morales, Tony H.
; TITLE OF INVENTION: Parasitic Helminth Cuticulin Proteins, Nucleic Acid
; FILE REFERENCE: Hw-8
; CURRENT APPLICATION NUMBER: US/09/323,427
; CURRENT FILING DATE: 1999-06-01
; EARLIER APPLICATION NUMBER: 60/087,435

```

```

; EARLIER FILING DATE: 1998-06-01
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 1161
; TYPE: DNA
; ORGANISM: Dicrofilaria immitis
US-09-323-427-5

Query Match      100.0%; Score 1161; DB 4; Length 1161;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGATGATGCTTATGCTTATGCTTCTGATGATGATGATGATGATGATGATGATGATGAT 60
Db 1161 ATGATGATGCTTATGCTTATGCTTCTGATGATGATGATGATGATGATGATGATGATGAT 1102
Qy 61 GTTGACAAATGCTGTCGAAAGTGGAGCCGGAATTTGAATGAGTACCACTTAATTAACATC 120
Db 1101 GTTGACAAATGCTGTCGAAAGTGGAGCCGGAATTTGAATGAGTACCACTTAATTAACATC 1042
Qy 121 AATTTAATACAGTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180
Db 1041 AATTTAATACAGTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 982
Qy 181 GAAGGTGCGGTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
Db 981 GAAGGTGCGGTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 922
Qy 241 TCATGCAATGTCGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
Db 921 TCATGCAATGTCGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 862
Qy 301 GTTGTATTCGTTTATCATCATTAATTTGTTACCAAGTTGATGCTGATATGATGATGAT 360
Db 921 GTTGTATTCGTTTATCATCATTAATTTGTTACCAAGTTGATGCTGATATGATGATGAT 802
Qy 361 TGCCTTTACATGAGAGCTGATTAACAGATTTAGTACAGATTTGAGTATCTGAATACCA 420
Db 801 TGCCTTTACATGAGAGCTGATTAACAGATTTAGTACAGATTTGAGTATCTGAATACCA 742
Qy 421 ACTGCTTTCAACATCAATATGTCGCCGATGCGGATGCTGCTGCTGCTGCTGCTGCTGCT 480
Db 741 ACTGCTTTCAACATCAATATGTCGCCGATGCGGATGCTGCTGCTGCTGCTGCTGCTGCT 682
Qy 481 GGACCAACCGGTCAACCGATTTGCTATCATTTGCTGACGAGTATTAATGATTAATG 540
Db 681 GGACCAACCGGTCAACCGATTTGCTATCATTTGCTGACGAGTATTAATGATTAATG 622
Qy 541 ACATGCGATTTGAAACCGTGTATGATCTTCTGCGGCTTGCATTTCTGCTTGTGAT 600
Db 621 ACATGCGATTTGAAACCGTGTATGATCTTCTGCGGCTTGCATTTCTGCTTGTGAT 562
Qy 601 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660
Db 561 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 502
Qy 661 TTGCTAATTAATTTGGAATATCAACAGATTTAATGCGTGGCAGAAAGCTACGTATAC 720
Db 501 TTGCTAATTAATTTGGAATATCAACAGATTTAATGCGTGGCAGAAAGCTACGTATAC 442
Qy 721 AAATATGCGGATGATCAACGCTTTCTATCAATGCGGATGATGATGATGATGATGATGATGAT 780
Db 441 AAATATGCGGATGATCAACGCTTTCTATCAATGCGGATGATGATGATGATGATGATGATGAT 382
Qy 781 CCAATATGCGAATGCTTTCGACACACATTTGATGATGATGATGATGATGATGATGATGATGAT 840
Db 381 CCAATATGCGAATGCTTTCGACACACATTTGATGATGATGATGATGATGATGATGATGATGAT 322
Qy 841 ACAGGTGTCGCCGACGAAACCTGCTGACGCTGCGCAACTGCTGCTGCTGCTGCTGCTGCT 900
Db 321 ACAGGTGTCGCCGACGAAACCTGCTGACGCTGCGCAACTGCTGCTGCTGCTGCTGCTGCT 262

```

Qy	TCGCGCAACCCGGAGATATCATTTGATGTCACGACATGATATCAACACCCCTGAAATTAGC	960
Db	TCGCGCAACCCGGAGATATCATTTGATGTCACGACATGATATCAACACCCCTGAAATTAGC	202
Qy	GATGATATATCAAGCTTTGGCCAGTTGGATTGATTACGTCACACCGTGCACTTCTGCACATATATGA	1020
Db	GATGATATATCAAGTTTGGCCAGTTGGATTGATTACGTCACACCGTGCACTTCTGCACATATATGA	142
Qy	CAACCTGTAATACTTCTGCTGCATGACAAATAATGGAATCTGCATCTACCACTTTGGCTTCTCA	1080
Db	CAACCTGTAATACTTCTGCTGCATGACAAATAATGGAATCTGCATCTACCACTTTGGCTTCTCA	82
Qy	ATGTTTATGGGTTTAAACATTCGATTCGATTCGGCCGTCATTATTAACATTTTCGTTTAA	1140
Db	ATGTTTATGGGTTTAAACATTCGATTCGATTCGGCCGTCATTATTAACATTTTCGTTTAA	22
Qy	TTTCGTCCAATCAGAAGCA	1161
Db	TTTCGTCCAATCAGAAGCA 1	

```

RESULT 3
US-09-812-642-3
: Sequence 3, Application US/09812642
: Patent No. 6368600
: GENERAL INFORMATION:
: APPLICANT: Chandrashekar, Ramaswamy
: APPLICANT: Morales, Tony H.
: TITLE OF INVENTION: Parasitic Helminth Cuticlin Proteins, Nucleic Acid
: FILE REFERENCE: HW-8
: CURRENT APPLICATION NUMBER: US/09/812,642
: CURRENT FILING DATE: 2001-03-20
: PRIOR APPLICATION NUMBER: 09/322,427
: PRIOR FILING DATE: 1999-06-01
: NUMBER OF SEQ ID NOS: 18
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 3
: LENGTH: 1161
: TYPE: DNA
: ORGANISM: Dirofilaria immitis
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (1)..(1161)
US-09-812-642-3

```

Query Match	100.0%	Score 1161;	DB 4;	Length 1161;
Best Local Similarly	100.0%;	Pred. No. 0;		
Matches 1161; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

QY	1	ATTNAGTATGCTTTATTTGCTTCTGTGACACACTTTTGCATTTGCTTTATTTGCAATTCG	60
Dp	1	ATGATGATTCCTCTTTATTTCTTTCTGTACTACACTTTATGCAATTTGCTTTATTTGCAATTCG	60
QY	61	GTTGACAAATGCTGTGCGAAGGTGAGCCGAAATTTGAATGTGACCACCTTCAATTAACATC	120
Dp	61	GTTGACAAATGCTGTGCGAAGGTGAGCCGAAATTTGAATGTGACCACCTTCAATTAACATC	120
QY	121	AATTTTAAATACCTAAATGCTTTGGAAGACATTTTATGGAAGCTCTTTATGATCAA	180
Dp	121	AATTTTAAATACCTAAATGCTTTGGAAGACATTTTATGGAAGCTCTTTATGATCAA	180
QY	181	GAAAGTTCCGTAATGATGAAGGTGACGCTCAAGTTCCGGAATTTCACTTCCATTTGAT	240
Dp	181	GAAAGTTCCGTAATGATGAAGGTGACGCTCAAGTTCCGGAATTTCACTTCCATTTGAT	240
QY	241	TCATGCAATGTTGGCGGACACGATCTCTGATCCACAGTGTATTTTGTAAACAACACT	300
Dp	241	TCATGCAATGTTGGCGGACACGATCTCTGATCCACAGTGTATTTTGTAAACAACACT	300
QY	301	GTTGTCATTTGTTTCATCCATTTATTTGTAAACAAAGTTGATGCTGCATATCGAGTACAA	360
Dp	301	GTTGTCATTTGTTTCATCCATTTATTTGTAAACAAAGTTGATGCTGCATATCGAGTACAA	360

QY	361	TGCTTTTACATGAGAGCTGATATAAACAAGTTAGTGGCAACAAATTGAGGTATCTGAAATCA	420
Db	361	TGCTTTTACATGAGAGAGCTGATATAAACAAGTTAGTGGCAACAAATTGAGGTATCTGAAATCA	420
QY	421	ACTGCTTTTCAAACTCAAAATTGTCCGATGCCAGTATGCGTATTAAGAAATTTTGGATGCT	480
Db	421	ACTGCTTTTCAAACTCAAAATTGTCCGATGCCAGTATGCGTATTAAGAAATTTTGGATGCT	480
QY	481	GGACCAACCGGTCACACGATTCCAATTTGCTATCATTTGCTACGCCAGTTATCATATAATGG	540
Db	481	GGACCAACCGGTCACACGATTCCAATTTGCTATCATTTGCTACGCCAGTTATCATATAATGG	540
QY	541	ACATGCGATTTCTGAACCCGTTGATACTTTCTGCGGGGTTGTCCATTCTGCTTGTGAT	600
Db	541	ACATGCGATTTCTGAACCCGTTGATACTTTCTGCGGGGTTGTCCATTCTGCTTGTGAT	600
QY	601	GATGTTAACGGTGATTTACTGTGGAAATTTCTAAATAGCTGATGATGATGCTTGATATAAT	660
Db	601	GATGTTAACGGTGATTTACTGTGGAAATTTCTAAATAGCTGATGATGATGCTTGATATAAT	660
QY	661	TTTGCTAAATAATTTTGGAAATATCCACAGATTTTAATGGCTGGCCAGAAGAGCTACGTAATAC	720
Db	661	TTTGCTAAATAATTTTGGAAATATCCACAGATTTTAATGGCTGGCCAGAAGAGCTACGTAATAC	720
QY	721	AAATATGCGGATGCGATACAGCGTTTCTCTATCAATGCGACATACAGATTAATCAATTAAGAA	780
Db	721	AAATATGCGGATGCGATACAGCGTTTCTCTATCAATGCGACATACAGATTAATCAATTAAGAA	780
QY	781	CCAAATAGCGAATGTGTTTCCACACCAATGTTTGAAACCCACAAGAGATTCGGAGCTGTATAA	840
Db	781	CCAAATAGCGAATGTGTTTCCACACCAATGTTTGAAACCCACAAGAGATTCGGAGCTGTATAA	840
QY	841	ACAGGTGCTGCCGAGCAAAACCTGCTGCAGCTGGCCAACTTCGTTTACTCAAGAAAAGA	900
Db	841	ACAGGTGCTGCCGAGCAAAACCTGCTGCAGCTGGCCAACTTCGTTTACTCAAGAAAAGA	900
QY	901	TCTCAGAGACCGGAGATTAATCATTTGATGTACAGAACTGATCAACACCCCTGGAATTTAGC	960
Db	901	TCTCAGAGACCGGAGATTAATCATTTGATGTACAGAACTGATCAACACCCCTGGAATTTAGC	960
QY	961	GATGATTAATCAAGCTTTGCCAGTTGATTTACGTACACCGTGCACTTCTGCAACATATAGGA	1020
Db	961	GATGATTAATCAAGCTTTGCCAGTTGATTTACGTACACCGTGCACTTCTGCAACATATAGGA	1020
QY	1021	CAACCTGTAAATACTTGCTGTAGTACAAATAGGAATCTGATGTACACATTTGGCTTCCA	1080
Db	1021	CAACCTGTAAATACTTGCTGTAGTACAAATAGGAATCTGATGTACACATTTGGCTTCCA	1080
QY	1081	ATGTTTATGGGTTTAAACATTTGATTTGATGGCCGTCATTTATTAACAATTCGTTTAAA	1140
Db	1081	ATGTTTATGGGTTTAAACATTTGATTTGATGGCCGTCATTTATTAACAATTCGTTTAAA	1140
QY	1141	TTTGCTCCAATTCGAAGGCA	1161
Db	1141	TTTGCTCCAATTCGAAGGCA	1161

```

? RESULT 4
? US-09-812-642-5/C
? Sequence 5, Application US/09812642
? Patent No. 6368600
? GENERAL INFORMATION:
? APPLICANT: Chandrashekar, Ramaswamy
? APPLICANT: Morales, Tony H.
? TITLE OF INVENTION: Parasitic Helminth Cuticlin Proteins, Nucleic Acid
? TITLE OF INVENTION: Molecules, and Uses Thereof
? FILE REFERENCE: HW-8
? CURRENT APPLICATION NUMBER: US/09/812,642
? CURRENT FILING DATE: 2001-03-20
? PRIOR APPLICATION NUMBER: 09/323,427
? PRIOR FILING DATE: 1999-06-01
? NUMBER OF SEQ ID NOS: 18
?

```

SOFTWARE: Patentin Ver. 2.0
 SEQ ID NO 5
 LENGTH: 1161
 TYPE: DNA
 ORGANISM: *Dirofilaria immitis*
 US-09-812-642-5

Query Match 100.0%; Score 1161; DB 4; Length 1161;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 ATGATGATGCTCTTATGCTTCTGTACTACCTTATGATGCTTATGATTCGG 60
    |||
DB 1161 ATGATGATGCTCTTATGCTTCTGTACTACCTTATGATGCTTATGATTCGG 1102
QY 61 GTTACATGCTGTCGAAAGTGAGCCAGAAATTAATGTGACCACTTCAATACATC 120
    |||
DB 1101 GTTACATGCTGTCGAAAGTGAGCCAGAAATTAATGTGACCACTTCAATACATC 1042
QY 121 AATTTAATACAGTAAATGATGATGAGACATGTTATGGAAGGCTTTATGATCAA 180
    |||
DB 1041 AATTTAATACAGTAAATGATGATGAGACATGTTATGGAAGGCTTTATGATCAA 982
QY 181 GAAGGTGCGGTAAATGATGAGGTGAGCGTCAAGTTCCGGAAATTCATTCATTTGAT 240
    |||
DB 981 GAAGGTGCGGTAAATGATGAGGTGAGCGTCAAGTTCCGGAAATTCATTCATTTGAT 922
QY 241 TCATGCAATGTGCGGTACACATCTCTGAATCCACGTGTATTTTGTACAAACACT 300
    |||
DB 921 TCATGCAATGTGCGGTACACATCTCTGAATCCACGTGTATTTTGTACAAACACT 862
QY 301 GTTGCTATTCGTTTCATTCATTTGTTACCAAGTTGATGCTGATTCGATACAGTACA 360
    |||
DB 861 GTTGCTATTCGTTTCATTCATTTGTTACCAAGTTGATGCTGATTCGATACAGTACA 802
QY 361 TCGTTTACATGAGAGCTGATTAACAGAGTATGACAGATGAGTATGTAATACACA 420
    |||
DB 801 TCGTTTACATGAGAGCTGATTAACAGAGTATGAGTATGAGTATGTAATACACA 742
QY 421 ACTGCTTTCAACTCAATGTGCGGATGCCGATGCCGTTATGAAATTTGGATGCT 480
    |||
DB 741 ACTGCTTTCAACTCAATGTGCGGATGCCGATGCCGTTATGAAATTTGGATGCT 682
QY 481 GGACCAACGGGTCAACAGTTCATTTGCTATGCTGACGACATTTATCATTAATG 540
    |||
DB 681 GGACCAACGGGTCAACAGTTCATTTGCTATGCTGACGACATTTATCATTAATG 622
QY 541 ACATGCAATGTGAAACCGTGTATCTTCTGCGGTTGCTCATCTGCTTTGCGAT 600
    |||
DB 621 ACATGCAATGTGAAACCGTGTATCTTCTGCGGTTGCTCATCTGCTTTGCGAT 562
QY 601 GATGTAACGGGTATCTGTGAAATTTCTAATGCTGATGATGCTCTTGTAAATAT 660
    |||
DB 561 GATGTAACGGGTATCTGTGAAATTTCTAATGCTGATGATGCTCTTGTAAATAT 502
QY 661 TTGCTAAATATTTGGAATATCCAAAGATTAATGGTGGCCAAAGACCTCAACGATAC 720
    |||
DB 501 TTGCTAAATATTTGGAATATCCAAAGATTAATGGTGGCCAAAGACCTCAACGATAC 442
QY 721 AAATATGCGGATGATCACAGCTTTTATCAATGCGCAGATCAGATTAATCAATTAAGAA 780
    |||
DB 441 AAATATGCGGATGATCACAGCTTTTATCAATGCGCAGATCAGATTAATCAATTAAGAA 382
QY 781 CCAAAATAGGAAATGTGTCGACCAATGTTCAAGAACCAAGAGATGCGACCTGTAA 840
    |||
DB 381 CCAAAATAGGAAATGTGTCGACCAATGTTCAAGAACCAAGAGATGCGACCTGTAA 322
QY 841 ACAGGTGTCGCGCAGCAAAACCTCTGCGAGCTGCGCAACTGTTTCTCAAGAAAGA 900
    |||
DB 321 ACAGGTGTCGCGCAGCAAAACCTCTGCGAGCTGCGCAACTGTTTCTCAAGAAAGA 262
QY 901 TCTGAGAAACCGGAGAAATATCATTTGATGACAACTGATACACCCCTGAAATTAGC 960
    |||

```

```

DB 261 TCTGAGAACCGGAGAAATATCATTTGATGTACGAACCTGATATCAACCCCTGAAATAGC 202
QY 961 GATGATAATCAAGCTTTGCCAGTTGATTTACGTCACCGGTGACTTTCGAACATATGGA 1020
    |||
DB 201 GATGATAATCAAGCTTTGCCAGTTGATTTACGTCACCGGTGACTTTCGAACATATGGA 142
QY 1021 CAACCTGTAAATCTTCTGTCGATACAAATGGAATCTGATGTCACCATTTGGCTTCYCA 1080
    |||
DB 141 CAACCTGTAAATCTTCTGTCGATACAAATGGAATCTGATGTCACCATTTGGCTTCYCA 82
QY 1081 ATGTTTATGAGTTTAAACATTCATGATGCTGCGCTCAATTTATACATTTGGTTAAA 1140
    |||
DB 81 ATGTTTATGAGTTTAAACATTCATGATGCTGCGCTCAATTTATACATTTGGTTAAA 22
QY 1141 TTTCGTCCAAATCAGAAGGCA 1161
    |||
DB 21 TTTCGTCCAAATCAGAAGGCA 1

```

RESULT 5.
 US-09-323-427-1
 Sequence 1, Application US/09323427
 Patent No. 6248329
 GENERAL INFORMATION:
 APPLICANT: Chandrasekar, Ramaswamy
 APPLICANT: Morales, Tony H.
 TITLE OF INVENTION: Parasitic Helminth Cuticulin Proteins, Nucleic Acid
 FILE REFERENCE: MW-8
 CURRENT APPLICATION NUMBER: US/09/323,427
 EARLIER APPLICATION NUMBER: 60/087,435
 EARLIER FILING DATE: 1998-06-01
 NUMBER OF SEQ ID NOS: 18
 SOFTWARE: Patentin Ver. 2.0
 SEQ ID NO 1
 LENGTH: 1779
 TYPE: DNA
 ORGANISM: *Dirofilaria immitis*
 US-09-323-427-1

Query Match 100.0%; Score 1161; DB 4; Length 1779;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 ATGATGATTCGCTTATGCTTCTGTACTACCTTATGATGCTTATGATTCGG 60
    |||
DB 167 ATGATGATTCGCTTATGCTTCTGTACTACCTTATGATGCTTATGATTCGG 226
QY 61 GTTGACAAATGCTGAGAGGTGAGCCAGAAATGGAATGTGACCAACTTCAATTAACATC 120
    |||
DB 227 GTTGACAAATGCTGAGAGGTGAGCCAGAAATGGAATGTGACCAACTTCAATTAACATC 286
QY 121 AATTTAATACAGTAAATGATGATGAGACATGTTATGCAAGGCTTTATGATACAA 180
    |||
DB 287 AATTTAATACAGTAAATGATGATGAGACATGTTATGCAAGGCTTTATGATACAA 346
QY 181 GAAGGTGCGGTAAATGATGAGGTGAGCTCAAGTTGCCGAATTTCACTTCCATTTGAT 240
    |||
DB 347 GAAGGTGCGGTAAATGATGAGGTGAGCTCAAGTTGCCGAATTTCACTTCCATTTGAT 406
QY 241 TCATGCAATGTGCGGTACACAGATCTCTGAATTCACAGCTGTATTTTGTAAACAACACT 300
    |||
DB 407 TCATGCAATGTGCGGTACACAGATCTCTGAATTCACAGCTGTATTTTGTAAACAACACT 466
QY 301 GTTGCTATTCGTTTCATTCATTTGTTACCAAGTTGATGCTGATTCGATACAGTACAA 360
    |||
DB 467 GTTGCTATTCGTTTCATTCATTTGTTACCAAGTTGATGCTGATTCGATACAGTACAA 526
QY 361 TCGTTTACATGAGAGCTGATTAACAGATGATGTCACAGATTTGAGATTTCTGAATACCA 420
    |||
DB 527 TCGTTTACATGAGAGCTGATTAACAGATGATGTCACAGATTTGAGATTTCTGAATACCA 586
    |||

```

QY 421 ACTGCTTTTCAAACTCAAAATTTGCCGATGCGACATATGCCGTTATGAAATTTTGGATGT 480
|||||
Db 587 ACTGCTTTTCAAACTCAAAATTTGCCGATGCGACATATGCCGTTATGAAATTTTGGATGT 646
|||||
QY 481 GGACCAACCGGTCACACAGTTTCAATTTGCTATCTATGTCAGCCAGTTTATCATTAATGG 540
|||||
Db 647 GGACCAACCGGTCACACAGTTTCAATTTGCTATCTATGTCAGCCAGTTTATCATTAATGG 706
|||||
QY 541 ACATGCGATCTGAACCCGTTGATCTTTGCGCGGTTGCGCATTTCTGCTTGTGAT 600
|||||
Db 707 ACATGCGATCTGAACCCGTTGATCTTTGCGCGGTTGCGCATTTCTGCTTGTGAT 766
|||||
QY 601 GATGTAACGGTGTACTGTGGAATTTCTAAATCTGATGATGTGCTTGTGATAATAT 660
|||||
Db 767 GATGTAACGGTGTACTGTGGAATTTCTAAATCTGATGATGTGCTTGTGATAATAT 826
|||||
QY 661 TTGCTAAATTAATTTGGAATTTCCACAGATTTAATGCTGCGCAAGAGCTCACGTATAC 720
|||||
Db 827 TTGCTAAATTAATTTGGAATTTCCACAGATTTAATGCTGCGCAAGAGCTCACGTATAC 886
|||||
QY 721 AAATATGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 780
|||||
Db 887 AAATATGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 946
|||||
QY 781 CCAATATGCGAATGTTTGCACACAAATGTTCAAGACCAACAGATTCGAGCTGTTAAA 840
|||||
Db 947 CCAATATGCGAATGTTTGCACACAAATGTTCAAGACCAACAGATTCGAGCTGTTAAA 1006
|||||
QY 841 ACAGGTGTCGCGCAGCAAAACCTGCTGACGCTGCGCACTTGTCTTACTCAAGAAAAGA 900
|||||
Db 1007 ACAGGTGTCGCGCAGCAAAACCTGCTGACGCTGCGCACTTGTCTTACTCAAGAAAAGA 1066
|||||
QY 901 TCTGCAACACCGGATATCATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 960
|||||
Db 1067 TCTGCAACACCGGATATCATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1126
|||||
QY 961 GATGATATCAAGCTTTGCCAGTTGATTTTACGTCACCGTGCACCTTCTGCAACATTAATGA 1020
|||||
Db 1127 GATGATATCAAGCTTTGCCAGTTGATTTTACGTCACCGTGCACCTTCTGCAACATTAATGA 1186
|||||
QY 1021 CAACCTGTAATCTTGTGCTGCAAGTACAAATGGAATCTGCAATGTCACATTTGGCTTCCA 1080
|||||
Db 1187 CAACCTGTAATCTTGTGCTGCAAGTACAAATGGAATCTGCAATGTCACATTTGGCTTCCA 1246
|||||
QY 1081 ATGTTTATGCGTTTAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140
|||||
Db 1247 ATGTTTATGCGTTTAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1306
|||||
QY 1141 TTTGCTCAAAATCAGAAAGCA 1161
|||||
Db 1307 TTTGCTCAAAATCAGAAAGCA 1327
|||||
RESULT 6
US-09-323-427-2/c
; Sequence 2, Application US/09323427
; Patent No. 6248329
; GENERAL INFORMATION:
; APPLICANT: Chandrashekar, Ramaswamy
; APPLICANT: Morales, Tony H.
; TITLE OF INVENTION: Parasitic Helminth Cuticulin Proteins, Nucleic Acid
; FILE REFERENCE: HW-8
; CURRENT APPLICATION NUMBER: US/09/323,427
; CURRENT FILING DATE: 1999-06-01
; EARLIER APPLICATION NUMBER: 60/087,435
; EARLIER FILING DATE: 1998-06-01
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1779
; TYPE: DNA
; ORGANISM: Dicrofilaria immitis

US-09-323-427-2
Query Match 100.0%; Score 1161; DB 4; Length 1779;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1161; Conserved 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGATGATGCTGCTTAAATGCTTTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 60
|||||
Db 1613 ATGATGATGCTGCTTAAATGCTTTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1554
|||||
QY 61 GTTGACAATGCTGTGCAAGGTGACCCAGAAATTAATGTCAGCAACTTCAATTAATATC 120
|||||
Db 1553 GTTGACAATGCTGTGCAAGGTGACCCAGAAATTAATGTCAGCAACTTCAATTAATATC 1494
|||||
QY 121 AATTTTAAATACGTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180
|||||
Db 1493 AATTTTAAATACGTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1434
|||||
QY 181 GAAAGTTGCGCTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
|||||
Db 1433 GAAAGTTGCGCTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1374
|||||
QY 241 TCATGCAATGTTGCGCGTACACGATCTGTAATCCAGTGTATTTTGTAAACAACACT 300
|||||
Db 1373 TCATGCAATGTTGCGCGTACACGATCTGTAATCCAGTGTATTTTGTAAACAACACT 1314
|||||
QY 301 GTTGCAATTTGCTTCAATCCATTTTGTGTCACAAAGTTGATGCTATGATGATGATGATGATGATGAT 360
|||||
Db 1313 GTTGCAATTTGCTTCAATCCATTTTGTGTCACAAAGTTGATGCTATGATGATGATGATGATGATGAT 1254
|||||
QY 361 TGTCTTTCATGGAAGCTGATTAACAAAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 420
|||||
Db 1253 TGTCTTTCATGGAAGCTGATTAACAAAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 1194
|||||
QY 421 ACTGCTTTTCAAACTCAAAATTTGCCGATGCGACATGTCGCTTATGAAATTTTGGATGCT 480
|||||
Db 1193 ACTGCTTTTCAAACTCAAAATTTGCCGATGCGACATGTCGCTTATGAAATTTTGGATGCT 1134
|||||
QY 481 GGACCAACCGGTCACACAGTTTCAATTTGCTATGATGATGATGATGATGATGATGATGATGATGATGAT 540
|||||
Db 1133 GGACCAACCGGTCACACAGTTTCAATTTGCTATGATGATGATGATGATGATGATGATGATGATGATGAT 1074
|||||
QY 541 ACATGCGATTTGGAACCCGTTGATCTTTCGCGGTTGTCATTTCTGCTTGTGTCAT 600
|||||
Db 1073 ACATGCGATTTGGAACCCGTTGATCTTTCGCGGTTGTCATTTCTGCTTGTGTCAT 1014
|||||
QY 601 GATGTAACCGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660
|||||
Db 1013 GATGTAACCGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 954
|||||
QY 661 TTTGCTAAATTAATTTGGAATTTCCACAGATTTAATGCTGCGCAAGAGCTCACGTATAC 720
|||||
Db 953 TTTGCTAAATTAATTTGGAATTTCCACAGATTTAATGCTGCGCAAGAGCTCACGTATAC 894
|||||
QY 721 AAATATGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 780
|||||
Db 893 AAATATGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 834
|||||
QY 781 CCAATATGCGAATGTTTGCACACAAATGTTCAAGACCAAGATTTGGAGCTTTAAA 840
|||||
Db 833 CCAATATGCGAATGTTTGCACACAAATGTTCAAGACCAAGATTTGGAGCTTTAAA 774
|||||
QY 841 ACAGGTGTCGCGCAGCAAAACCTGCTGACGCTGCGCACTTGTCTTACTCAAGAAAAGA 900
|||||
Db 773 ACAGGTGTCGCGCAGCAAAACCTGCTGACGCTGCGCACTTGTCTTACTCAAGAAAAGA 714
|||||
QY 901 TCTGCAAGACCGGATATCATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 960
|||||
Db 713 TCTGCAAGACCGGATATCATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 654
|||||
QY 961 GATGATATCAAGCTTTGCCAGTTGATTTTACGTCACCGTGCACCTTCTGCAACATTAATGA 1020
|||||
Db 653 GATGATATCAAGCTTTGCCAGTTGATTTTACGTCACCGTGCACCTTCTGCAACATTAATGA 594
|||||

QY 1021 CAACCTGTAATCTTGTGACGATCAAAATGGAATCTGATGACCATTTGGCTTCGA 1080
|
Db 593 CAACCTGTAATCTTGTGACGATCAAAATGGAATCTGATGACCATTTGGCTTCGA 534
|
QY 1081 ATGTTATGGGTTTAAGCATTTGATTCGCGGTCATATTATACCATTTGGTTTAA 1140
|
Db 533 ATGTTATGGGTTTAAGCATTTGATTCGCGGTCATATTATACCATTTGGTTTAA 474
|
QY 1141 TTGCTCCAATCAGAAGCA 1161
|
Db 473 TTGCTCCAATCAGAAGCA 453
|
RESULT 7
US-09-812-642-1
; Sequence 1, Application US/09812642
; Patent No. 6368600
; GENERAL INFORMATION:
; APPLICANT: Chandrashekar, Ramaswamy
; TITLE OF INVENTION: Parasitic Helminth Cuticlin Proteins, Nucleic Acid
; FILE REFERENCE: HW-8
; CURRENT APPLICATION NUMBER: US/09/812,642
; PRIOR FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: 09/323,427
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1779
; TYPE: DNA
; ORGANISM: Dirofilaria immitis
US-09-812-642-1

Query Match 100.0%; Score 1161; DB 4; Length 1779;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGATGATTCGCTTATTCGTTTCTGTACTACATTAATGTCATTTGCTTATTCGATTCG 60
|
Db 167 ATGATGATTCGCTTATTCGTTTCTGTACTACATTAATGTCATTTGCTTATTCGATTCG 226
|
QY 61 GTTGACAATGCTGTCGAGGTGAGCCAGAAATTTGAATGTGAGCCCACTTCATTAACATC 120
|
Db 227 GTTGACAATGCTGTCGAGGTGAGCCAGAAATTTGAATGTGAGCCCACTTCATTAACATC 286
|
QY 121 AATTTTAAATACAGTAATGTCATTCGAGACATGTTTATGTAAGTCTTTATGATCA 180
|
Db 287 AATTTTAAATACAGTAATGTCATTCGAGACATGTTTATGTAAGTCTTTATGATCA 346
|
QY 181 GAAGGTGCGCTTAATGATGAGGTGAGCTCAAGTTCGCGAATTTCACTTCATTTGAT 240
|
Db 347 GAAGGTGCGCTTAATGATGAGGTGAGCTCAAGTTCGCGAATTTCACTTCATTTGAT 406
|
QY 241 TCATGCAATGTTGCGCGTACAGATCTCGATCCACGCGTATTTTGTATTAACAACACT 300
|
Db 407 TCATGCAATGTTGCGCGTACAGATCTCGATCCACGCGTATTTTGTATTAACAACACT 466
|
QY 301 GTTGCAATTTGTTTCAATCAATTTATTTGTACCAAGTTGATCGCATTCGATGAGTAC 360
|
Db 467 GTTGCAATTTGTTTCAATCAATTTATTTGTACCAAGTTGATCGCATTCGATGAGTAC 526
|
QY 361 TGCCTTTACATGAGAGCTGATTAACAGTATGTCACAGATTTAGATTCGAATTCACA 420
|
Db 527 TGCCTTTACATGAGAGCTGATTAACAGTATGTCACAGATTTAGATTCGAATTCACA 586
|
QY 421 ACTGCTTTTCAACGCAATGTTCCGATGGCAGATGCCCTTATGAATTTTGGATGCT 480
|
Db 587 ACTGCTTTTCAACGCAATGTTCCGATGGCAGATGCCCTTATGAATTTTGGATGCT 646
|
QY 481 GGACCAACCGGTCAACAGTTCATTTGCTATCATTTGCTGAGCCAGTTTATCATTAATG 540
|

|
Db 647 GGACCAACCGGTCAACAGTTCATTTGCTATCATTTGCTGAGCCAGTTTATCATTAATG 706
|
QY 541 ACATGCGATTTGTAACCGTGTATCTTTCGCGCGGTTCATCTCTGTTTGCAT 600
|
Db 707 ACATGCGATTTGTAACCGTGTATCTTTCGCGCGGTTCATCTCTGTTTGCAT 766
|
QY 601 GATGTAACCGGTGATCTGTAATCTTAATCTGTGATGATGCTTGTATTAAT 660
|
Db 767 GATGTAACCGGTGATCTGTAATCTTAATCTGTGATGATGCTTGTATTAAT 826
|
QY 661 TTGCTAAATTAATTTGGAATTCACAGATTTAATGGCTGGCCAAAGCTCAGTATAC 720
|
Db 827 TTGCTAAATTAATTTGGAATTCACAGATTTAATGGCTGGCCAAAGCTCAGTATAC 886
|
QY 721 AATATGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 780
|
Db 887 AATATGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 946
|
QY 781 CCAATATGCGAATGTTTCGACCAATGTTTCGACCAATGTTTCGACCAATGTTTCGAC 840
|
Db 947 CCAATATGCGAATGTTTCGACCAATGTTTCGACCAATGTTTCGACCAATGTTTCGAC 1006
|
QY 841 ACAGGTGTCGCCGACGCAAACTGCTGACGTCGCGCACTGCTTACTCAAGAAAGA 900
|
Db 1007 ACAGGTGTCGCCGACGCAAACTGCTGACGTCGCGCACTGCTTACTCAAGAAAGA 1066
|
QY 901 TCTGCAAGACCGGAGATATCATTTGATGATGATGATGATGATGATGATGATGATGAT 960
|
Db 1067 TCTGCAAGACCGGAGATATCATTTGATGATGATGATGATGATGATGATGATGATGAT 1126
|
QY 961 GATGATATCAACCTTGGCAGTTGATGATGATGATGATGATGATGATGATGATGATGAT 1020
|
Db 1127 GATGATATCAACCTTGGCAGTTGATGATGATGATGATGATGATGATGATGATGATGAT 1186
|
QY 1021 CAACCTGTAATCTTGTGACGATCAAAATGGAATCTGATGACCATTTGGCTTCGA 1080
|
Db 1187 CAACCTGTAATCTTGTGACGATCAAAATGGAATCTGATGACCATTTGGCTTCGA 1246
|
QY 1081 ATGTTATGGGTTTAAGCATTTGATTCGCGGTCATATTATACCATTTGGTTTAA 1140
|
Db 1247 ATGTTATGGGTTTAAGCATTTGATTCGCGGTCATATTATACCATTTGGTTTAA 1306
|
QY 1141 TTGCTCCAATCAGAAGCA 1161
|
Db 1307 TTGCTCCAATCAGAAGCA 1327
|
RESULT 8
US-09-812-642-2/c
; Sequence 2, Application US/09812642
; Patent No. 6368600
; GENERAL INFORMATION:
; APPLICANT: Chandrashekar, Ramaswamy
; TITLE OF INVENTION: Parasitic Helminth Cuticlin Proteins, Nucleic Acid
; FILE REFERENCE: HW-8
; CURRENT APPLICATION NUMBER: US/09/812,642
; PRIOR FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: 09/323,427
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1779
; TYPE: DNA
; ORGANISM: Dirofilaria immitis
US-09-812-642-2

Query Match 100.0%; Score 1161; DB 4; Length 1779;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1081	ATTTTANGGTTAAGCATGTGATGCGCGCTCATTTATACATTGGTTAA	1141	TTTGTCCAAATCAGAAGCA	1161
DB	533	ATGTTTATGAGTTTAAGCATGTGATGCGCGCTCATTTATACATTGGTTAA	473	TTTGTCCAAATCAGAAGCA	453
QY	1141	TTTGTCCAAATCAGAAGCA	1161	TTTGTCCAAATCAGAAGCA	1161
DB	473	TTTGTCCAAATCAGAAGCA	453	TTTGTCCAAATCAGAAGCA	453
<p>RESULT 9</p> <p>US-09-323-427-6</p> <p>Sequence 6, Application US/09323427</p> <p>Patent No. 6248329</p> <p>GENERAL INFORMATION:</p> <p>APPLICANT: Chandrasekar, Ramaswamy</p> <p>APPLICANT: Morales, Tony H.</p> <p>TITLE OF INVENTION: Parasitic Helminth Cuticlin Proteins, Nucleic Acid</p> <p>TITLE OF INVENTION: Molecules, and Uses Thereof</p> <p>FILE REFERENCE: HW-8</p> <p>CURRENT APPLICATION NUMBER: US/09/323,427</p> <p>CURRENT FILING DATE: 1999-06-01</p> <p>EARLIER APPLICATION NUMBER: 60/087,435</p> <p>EARLIER FILING DATE: 1998-06-01</p> <p>NUMBER OF SEQ ID NOS: 18</p> <p>SOFTWARE: PatentIn Ver. 2.0</p> <p>SEQ ID NO 6</p> <p>LENGTH: 1372</p> <p>TYPE: DNA</p> <p>ORGANISM: <i>Dirofilaria immitis</i></p> <p>US-09-323-427-6</p>					
<p>Query Match 43.4%; Score 504.2; DB 4; Length 1372;</p> <p>Best Local Similarity 76.9%; Pred. No. 3.7e-137;</p> <p>Matches 652; Conservative 0; Mismatches 193; Indels 3; Gaps</p>					
QY	1	ATGATGATTCGCTTATTGCTTCTGTACTACACTTATG-CATGTCTTATCGATCC	59		
DB	32	ATATACATTTATTCCTGGTATTCACATTTCTACTTAACCTACCTACATCATGATCCC	91		
QY	60	GGTTGACAAATGGTGTGCAAGGTGACGACAGAAATGATGTGACCACTCAATTAACAT	119		
DB	92	TATTTGACATGTGTGTCAGAGTGACACTGAAATGATGTGCGGACCTTCGATACAT	151		
QY	120	CAATTTAATACACGTATATGATGCAAGGACATGTTTATGTGAAAGGCTTATGATCA	179		
DB	152	CAATTTCAATCTGAAATATACATTTTGAAGGACAGCTATATGTAAGAAGCATATGATCA	211		
QY	180	AGAAGGTTCCTGATATGATGACGTGACGTCAGAGTTGCCGAATTTCACTTCATTTGA	239		
DB	212	GGATGATATGTCGTGACATGATATGACGCGCAGGTAGCTGAAATCAATTTGCAATGGA	271		
QY	240	TTTCATGCAATGTTCGGGTACAGATCTGCAATGCCAGCGGATTTTGTAAACAACAC	299		
DB	272	TTTCGTATATGTAAGCATGATCATCTTAAATCCGTGTGGTGTTTTGTAAACAACGT	331		
QY	300	TGTTGTATTTCTGTTATCATCATTTTGTTACCAAACTTGATGCTGATATGAGTACA	359		
DB	332	AGTTGTATTTATTTATTCATCCAAATTCGTTACAAAAATGATGACATATGCTATACA	391		
QY	360	ATGCTTTTACATGG-AAGCTGATTAACACAGTTAGTGC-ACAGATGAGTATCTGAATC	417		
DB	392	ATGCTTTTATATGGAAGAGCTGATTAAGACCGTTAGTACTGCTGTAAGATCTGAATG	451		
QY	418	ACAACGTCTTTTCAAACTCAAAATGTGCCATGGCACTATGCGCTTATGAAATTTGGAT	477		
DB	452	ACTACAGATTTCAAACTCAAAAGTGATGCAATGCCGTATATGATATGAGATTTGGAA	511		
QY	478	GGTAGACCAACCGGTCAACACAGTTCAATTTGTATCATCTTGTGTCAGCGATTTATCAA	537		
DB	512	GGTAGACCAACCGGTGACACCGCTTTGCAATTTGCAATGATTTGGAAGATCATGTATACAAA	571		
QY	538	TGACATGCGATTTCTGGAACCGTTGATATCTTCTGCGGGGTGTGCAATCTGCTTTGTC	597		

Db	572	TGACATGTGATTCAGAGACTACAGATACATCTCTGTCATTAGACATTCATGTGTTGTC	631
Qy	598	GATGATGCTAACCGTGATCTGCTGGAATTTCTAAATGCTGATGATGCTGCTTTGATAAA	657
Db	632	GATGATGGAAAAAGGTGATGATGCGTGAAGATTTCTGTAATGAAGAGATTTGCTTTGGACAAA	691
Qy	658	TATTTGCTAAATTAATTTGGATATATCCAACAGATTTAATGCTGCGCCAAAGCTCAGCTA	717
Db	692	TATTTACTCATTAATTTGGATATATATACAGATTTAATGGCTGGCCAAAGACTCATGTT	751
Qy	718	TACAAATATGCGGATCCGATCACAGCTTTTCTATCAATGCCAGATCAGTATTAACATTAAA	777
Db	752	TATATAATATGAGATCGATCGATCAACTTTCATATCAATGACCAAGATTTAATATAACAAATTTAAA	811
Qy	778	GAAACCAATATAGCAATGTGTGTGCGACCAATATGTCAGACACACAAGATTCGGAGCGT	837
Db	812	GAGCCACACTACGCAATGTCTTCGACCAACATATGCACACAGCCACACAGATTTGGTGCATTA	871
Qy	838	AAAAACAGG	845
Db	872	AAATCTGG	879

```

RESULT 10
US-09-323-427-7/c
? Sequence 7, Application US/09323427
? Patent No. 6248329
? GENERAL INFORMATION:
? APPLICANT: Chandrasekar, Ramaswamy
? APPLICANT: Morales, Tony H.
? TITLE OF INVENTION: Parasitic Helminth Cytolysin Proteins, Nucleic Acid
? TITLE OF INVENTION: Molecules, and Uses Thereof
? FILE REFERENCE: HW-8
? CURRENT APPLICATION NUMBER: US/09/323,427
? CURRENT FILING DATE: 1999-06-01
? EARLIER APPLICATION NUMBER: 60/087,435
? EARLIER FILING DATE: 1998-06-01
? NUMBER OF SEQ ID NOS: 18
? SOFTWARE: Patentln Ver. 2.0
? SEQ ID NO 7
? LENGTH: 1372
? TYPE: DNA
? ORGANISM: Dirofilaria immitis
? US-09-323-427-7

```

Query Match	43.48;	Score	504.2;	DB 4	Length	1372;	
Best Local Similarity	76.98;	Pred. No.	3.7e137;				
Matches 652; Conservative	0;	Mismatches	193;	Indels	3;	Gaps	3;

QY	1	ATGTCGATTCGCTCTATATGCTTTCTGTACTACACTATTG-CATTGCTCTATTGATTC	59
Db	1341	ATTATCATTTTATTCCTTGTATTCTCCACTTTTCTACCTGACCTACGATACATGATCC	1282
QY	60	GCTTGCACATGCTGTGGAAGGTGAGCCACAATAATGATGTGACCACTTCATTAACAT	119
Db	1281	TATTGACAAATGCTGTGGAAGGTGACCTTAAATAGATGTGCGCAGCTTGATTAACAT	1222
QY	120	CAATTTTAAATACAGTAATGCAATTCGCAAGACATGTTATGGAAGTCTTTATGATCA	179
Db	1221	CAATTTCAATACAGTAATGCAATTCGCAAGACATGTTATGGAAGTCTTTATGATCA	1152
QY	180	AGAAAGTTGCCGTAATGATGAAGGTGAGCGTCAAGTTGCCGAATTTCACTTCATTTGA	239
Db	1161	GGATGATGTCGTTACGATAGTAATGAGGAGCGAGGTGGAATCGAATTTGGCAATGA	1102
QY	240	TTTATGCAATGTTGCCGCTACACGATCTCTGAATCCACGCTGATTTTGTAAACAAC	299
Db	1101	TTTCGTAAATGTTGAACGATACGATCCCTTAATCTCTGTGTTTGTAAACAACGT	1042
QY	300	TGTTGTCATTTGCTTTCATCCATATTATTGTACAAAGTGAATGCTGATATCGATACA	359
Db	1041	AGTTGTCATTTTACATTTTACCAAAATTCGTTTCAAAAATAGTACGAGATTCGATACA	982

OY	360	ATGCTTTACATGCG - AACCTATAAAGACGTTAGTGC - ACAGATTGAGGATATCTGAATC	41.7
Db	961	ATGTTTATATATGGAAACGCTATAGAACCGCTTACTGCTGCTAGGATATCTGGAATG	92.2
OY	418	ACAACCTGCTTTTCAACCAACAATATGTCCCGATGCGCAGATGCGCTATGAAATTTTGAT	47.7
Db	921	ACTACAGCATTTCCAACATCAAGTGGTACCAATGCGCCGATGTGCATATGAGATTTTTGAA	86.6
OY	478	GGTGACCAACCCGTCACACCACTTCATTTGCTATCATTTGCTACGCCAGTTTATCATAAA	53.3
Db	861	GGTGACCAACCTGGTGACACCGTTCGATTTGCCAATGATTTGAGATTCATATATACAAA	80.2
OY	538	TGCAACCTCGATTTGGAACCGTGTGATACCTTTTGGCGCGGTTGCCATTCCTGCTTGTG	59.7
Db	801	TGGACATGTGATTCAGAGACTACAGATATCATTTCTGTGCTATTAGTACATTCATGTGTTG	74.2
OY	598	GATGATGCTAACGGTGATCTGTGGAAATTTCTAAATCTGATGAGTGTGCTCTTGATAAA	65.7
Db	741	GATGATGGAAGAGGTGATGCTAGTGGAAATTTCTGATATGAAGAGATGTGCTTTGGACAAA	68.2
OY	658	TATTTTGCTAATATTTTGGATATTCACCAAGATTTAATGGCTGGCCCAAGAGCTACGTA	71.7
Db	661	TATTTACTCATTAATTTTGGATATATTTACACATTTTAATGGCTGGCCCAAGAGCTACGTT	62.2
OY	718	TACAATATGCGGATCGATCCACAGCTTTTCTATCAATGCCAGATCACTATTAACATTTAA	77.7
Db	621	TATTAATATGCAAGATCGATCCAGACTTTACTATCAATGCAGATATTATATACAAATAAA	56.2
OY	778	GAACCAATATACGAATGTGTCGACCAACAATGTTCAACAACCAAGATTCGGAAGCTGT	83.7
Db	561	GAGCCACATATACGAATGCTCTCGACCAACAATGCAACAAGCCACAGGATTTGGTGCCATA	50.2
OY	838	AAAAACAGG 84.5	
Db	501	AAATCTGG 49.4	

```

RESULT 11
US-09-812-642-6
? Sequence 6, Application US/09812642
? Patent No. 6368600
? GENERAL INFORMATION:
? APPLICANT: Chandrashekar, Ramaswamy
? APPLICANT: Morales, Tony H.
? TITLE OF INVENTION: Parasitic Helminth Cuticlin Proteins, Nucleic Acid
? TITLE OF INVENTION: Molecules, and Uses Thereof
? FILE REFERENCE: HW-8
? CURRENT APPLICATION NUMBER: US/09/812,642
? CURRENT FILING DATE: 2001-03-20
? PRIOR APPLICATION NUMBER: 09/323,427
? PRIOR FILING DATE: 1999-06-01
? NUMBER OF SEQ ID NOS: 18
? SOFTWARE: PatentIn Ver. 2.0
? SEQ ID NO 6
? LENGTH: 1372
? TYPE: DNA
? ORGANISM: Dirofilaria immitis
US-09-812-642-6

```

Query Match	43.48;	Score 504.2;	DB 4;	Length 1372;
Best Local Similarity	76.98;	Pred. No. 3.7e137;		
Matches 652; Conservative	0;	Mismatches 193;	Indels 3;	Gaps 3;

OY	1	AAGTAGATTCGCTTAATTGCCTTTCGTACTACACTAATTG-CATTGCTTAATCGAATCC	59
Db	32	ATTATCATTTATCTCCTTGTTATTCCTAACTTTCTACTGCACTATCATCATGCC	91
OY	60	GGTGACAAGGGTGTGCAAGGTGAGCCAGAATAATGAATGGAGCACCTCAATAACAAT	119
Db	92	TATTGACAATGGTGTCCAGAGGTGAACCTGAATATGAATGTGGCCACCTTCGATAACAT	151
OY	120	CAATTTTAATACACGTAAATGATTCGAAAGACATGTTTATGTGAAGGTCTTAATGATCA	179

Db 152 CAATTCATCTAGAAAATCAATTTGAAGACACGATATATGTAAGAACTATATGATCA 211
180 AGAAGTTGCGCTAATGATGAGGTGAGCGTCAAGTTGCCGAATTCACATTCATTTGA 239
212 GGAATGATGTCGTTCAATGATGATGACGCGAGTACGCTGGAATTCGAATTTGGCAATGA 271
240 TTGATGCAATGTTGCGGTCACAGATCTGAAATCCACGCTGATATTTTGTAAACAAAC 299
272 TTGCTGAATGTTGAACGATCAGATCCTTAATCTCGTGTGTTTGTGTAACAACTGT 331
300 TGTGTCATTTGCTTTCATTCATTTATTTGTACCAAGTATGCTGATGAGTACA 359
332 AGTTGTCATTTACATTTTCAATTCGTTCAAAAATGATCGACAAATTCGTATACA 391
360 ATGCTTTTACATGAG -AAGCTGATATAAACAGTATAGTGC -ACAGATTTGAGTATCGAATC 417
392 ATGTTTTTATATGGAAGAGCTGATAGACCGTTAGTACTGCTCTTGAAGTATCTGAATG 451
418 ACAATGCTTTTCAAACTCAAAATTTGCCGATGCGATGCGTTATGAAATTTGGAT 477
452 ACTAGACATTTCCAAACTCAAGTGTACCAATGCCGTATGTGATGATGATTTTGGAA 511
478 GGTGACCAACCGGTGACACCAATTTGCTATCTATGCTGACGCACTTATATCAAAA 537
512 GGTGGACCAACTGCTGACCTGTTGCAATGATGATGAGATCATGATATACAAAA 571
538 TGGACATGCGATTTCTGAACCGTTGATACCTTCTGCGCGGTGCTGCTGCTTGTGC 597
572 TGGACATGCTATTCACAGACTACAGATACATCTGCTGATAGTACATTCATGCTGTG 631
598 GATGATGTAACGCTGATGATCTGGAATTTCTAATGCTGATGATGCTGCTTGTATAAA 657
632 GATGATGGAAGAGTGTGATGAGTGTGATCTGATGATGAAGAAGATGCTTTGGAAAA 691
658 TATTTGCTAATTAATTTGGAATTCACACAGATTTAATGCTGCGCAAGAGCTCAGTA 717
692 TATTTACTCAATTAATTTGGAATTAATTAATTAATTAATGCTGCGCAAGAGCTCATGTT 751
718 TACAATATGCGGATGATGATACAGCTTTTATCAATGCCAGATATATATACATTA 777
752 TATTAATATGCAATGATGATGAACTTACTATCAATGCCAGATTAATTAACAAATTA 811
778 GAACCAATATGCAATGCTGTTGACCAACCAATGTTCAAGACACAGAGTCTGAGCTGT 837
812 GAGCCACATAGCGAATGCTCTCGACCAATGACACAGACACAGAGATTTGGTCCATA 871
QY 838 AAAACAG 845
111 111
Db 872 AAATCTGG 879

RESULT 12
US-09-812-642-7/c
; Sequence 7, Application US/09812642
; Patent No. 6368600
; GENERAL INFORMATION:
; APPLICANT: Chandrashekar, Ramaswamy
; APPLICANT: Morales, Tony H.
; TITLE OF INVENTION: Parasitic Helminth Cuticlin Proteins, Nucleic Acid
; FILE REFERENCE: HM-8
; CURRENT APPLICATION NUMBER: US/09/812,642
; PRIORITY FILING DATE: 2001-03-20
; PRIORITY APPLICATION NUMBER: 09/323,427
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 7
; LENGTH: 1372
; TYPE: DNA
; ORGANISM: *Dirofilaria immitis*
US-09-812-642-7

Query Match 43.4%; Score 504.2; DB 4; Length 1372;
Best Local Similarity 76.9%; Pred. No. 3.7e-137;
Matches 652; Conservative 0; Mismatches 193; Indels 3; Gaps 3;

QY 1 ATGATGATTTGCTCTTATTTGCTTTCTGTACTACACTTATG -CATGTCCTTATGATTC 59
1341 ATTAATCATTTATTCCTTGTATTCCTTCTTCTACTTGTGACCTACGATCATCATGCC 1282
60 GGTGACAAATGCTGCAAGGTGACGCAAAATGAAATGAGGACCAATCAATTAACAT 119
1281 TATTTGCAATGATGCTGCAAGGTGACGCAAAATGAAATGAGGACCAATCAATTAACAT 1222
QY 120 CAATTTAATACAGTAATGATTCGAAAGACATGTTATGTAAGAGTCTTATGATCA 179
1221 CAATTTCAATACAGTAATGATTCGAAAGACATGTTATGTAAGAGTCTTATGATCA 1162
QY 180 AGAAGTGTCCGCTATATGATGAGTGTGACGCTCAAGTTGCCGAATTTCACTTCATTTGA 239
1161 GGAATGAAATGCTGCTCAATGATGATGAGGACGCGAGTGTGAAATTCGAATTTGGCAATGGA 1102
240 TTGATGCAATGTTGCGGTGACAGATCTGAAATCCACGCTGATTTTGTGAACAAAC 299
1101 TTGCTGTAATGTTGAAAGATCAGATCCTTAATCTCGTGTGTTTGTATACACTGT 1042
300 TGTGTCATTTGCTTTCATTCATTTTGTACCAAGTATGATGCTGATATCGAGTACA 359
1041 AGTTGCTATTTACATTTTATTCACAAATTCGTTACAAAATATGATGACATATCTATACA 982
QY 360 ATGCTTTTACATGAG -AAGCTGATATAAACAGTATAGTGC -ACAGATTTGAGTATCGAATC 417
981 ATGTTTTTATATGGAAGAGCTGATTAAGACCGTTAGTACTGTCCTTGAAGTATCGAATG 922
418 ACAATGCTTTTCAAACTCAAAATTTGCCGATGCTGATGCTGCTTATGAAATTTGGAT 477
921 ACTACACATTTCCAAACTCAAGTGTGACCAATGCGCTGATGATGAGATTTGGAA 862
478 GGTGACCAACCGGTGACACCAAGTTCATTTGCTGATGCTGCTGCTGCTGCTTATCAATA 537
861 GGTGACCAACCGGTGACACCAAGTTCATTTGCTGATGCTGCTGCTGCTGCTTATCAATA 802
QY 538 TGGACATGCGATTTCTGAACCGTTGATACCTTCTGCGCGGTGCTGCTTGTGCT 597
801 TGGACATGCTATTCACAGACTACAGATACATCTGCTGATAGTACATTCATGCTGTG 742
QY 598 GATGATGTAACGCTGATGATCTGGAATTTCTAATGCTGATGATGCTGCTTGTATAAA 657
741 GATGATGGAAGAGTGTGATGAGTGTGATGATGATGATGATGATGATGATGATGATG 682
658 TATTTGCTAATTAATTTGGAATTAATTAATTAATTAATTAATGCTGCGCAAGAGCTCAGTA 717
681 TATTTACTCAATTAATTTGGAATTAATTAATTAATTAATTAATGCTGCGCAAGAGCTCATGTT 742
QY 718 TACAATATGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 777
621 TATTAATATGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 742
QY 778 GAACCAATATGCAATGCTGTTGACCAACCAATGTTCAAGACACAGAGTCTGAGCTGT 837
561 GAGCCACATAGCGAATGCTCTCGACCAATGACACAGACACAGAGATTTGGTCCATA 502
QY 838 AAAACAG 845
111 111
Db 501 AAATCTGG 494

RESULT 13
US-09-323-427-16
; Sequence 16, Application US/09323427
; Patent No. 6248329
; GENERAL INFORMATION:
; APPLICANT: Chandrashekar, Ramaswamy
; APPLICANT: Morales, Tony H.
; TITLE OF INVENTION: Parasitic Helminth Cuticlin Proteins, Nucleic Acid

;; TITLE OF INVENTION: Molecules, and Uses Thereof
;; FILE REFERENCE: HW-8
;; CURRENT APPLICATION NUMBER: US/09/323,427
;; CURRENT FILING DATE: 1999-06-01
;; EARLIER APPLICATION NUMBER: 60/087,435
;; EARLIER FILING DATE: 1998-06-01
;; NUMBER OF SEQ ID NOS: 18
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO: 16
;; LENGTH: 892
;; TYPE: DNA
;; ORGANISM: Brugia malayi
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (158)..(892)
US-09-323-427-16

Query Match 39.7%; Score 461; DB 4; Length 892;
Best Local Similarity 78.1%; Pred. No. 1,1e-124;
Matches 554; Conservative 0; Mismatches 155; Indels 0; Gaps 0;

QY 21 TTTCGTACTACACTTATGCAATGCTTATTCGATTCGCGTTGACCAATGCTGCAAG 80
DB 184 TTTCGTACTATGATATGCAAGATTTATGCTATTCGAATGATGCTGCAAG 243
QY 81 TGAGCAGAAATTTGATGTCGACCACTTCATTAACATCAATTTTAATACAGTATGC 140
DB 244 TGAACCTGAAATTTGATGTCGACCACTTCATTAATTTTAATACAGTATGC 303
QY 141 ATTCCAGAGACATGTTATGTAAGGCTTTATGATGTAAGAGTGGCGTAATGATGA 200
DB 304 TTTTGAAGACATGATATGTAAGGATTTATGATGTAAGATGTCGATGATGA 363
QY 201 AGGTGAGAGCTCAAGTGGCGGAATTTCACTTCATTTGATTCAGCAATGTCGCGTAC 260
DB 364 AGGTGAGAGCTCAAGTGGCGGAATTTCACTTCATTTGATTCAGCAATGTCGCGTAC 423
QY 261 AGCATCTGCAATCCAGTGGTATTTTGTAAACAACAACGCTGTCATTTGCTTCATCC 320
DB 424 AGCTTCGTTAAATCCAGTGGTATTTGTCACAGCTGTCGTAATTAATGCTTCATCC 483
QY 321 ATTATTTGTTACCAAGTTGATGTCGATGATGATGATGATGATGATGATGATGATGAT 380
DB 484 ACAGTTTATCAAAAAGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 543
QY 381 TAAACAGTTAGTGCACAGATTTGATGATGATGATGATGATGATGATGATGATGATGAT 440
DB 544 TAAGACTGTTAGCACACAATTTGAAGTTCCGAATTTGACAAACCGTATTTGCTACCAATT 603
QY 441 TGTCCTGATGCCAGTATCCGTTATGAATTTTGGATGTCGACCAACCGGTCACAGT 500
DB 604 GGTACCAATGCTGCTGTAGATATGATGATGATGATGATGATGATGATGATGATGATGAT 663
QY 501 TCAATTTGCTATCATTTGTCAGCCGTTATCATTAATGATGATGATGATGATGATGATGAT 560
DB 664 CCAATGCTATCATTTGTCAGCCGTTATCATTAATGATGATGATGATGATGATGATGATGAT 723
QY 561 TGATACCTTTCGCGGTTGTCATTCCTGCTTGTGATGATGATGATGATGATGATGATGAT 620
DB 724 TGATACCTTTCGCGGTTGTCATTCCTGCTTGTGATGATGATGATGATGATGATGATGAT 783
QY 621 GGAATTTCTAATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 680
DB 784 TAAATTTAATTAATGAAGAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 843
QY 681 TCCAAAGATTTAATGCTGTCGCAAGAAGCTCAGCTATACAAATATGCG 729
DB 844 TCCAAAGATTTAATGCTGTCGCAAGAAGCTCAGCTATACAAATATGCG 892

RESULT 14
US-09-323-427-18/c
; Sequence 18, Application US/09323427

;; Patent No. 6248329
;; GENERAL INFORMATION:
;; APPLICANT: Chandrasekar, Ramaswamy
;; APPLICANT: Morales, Tony H.
;; TITLE OF INVENTION: Parasitic Helminth Cuticlin Proteins, Nucleic Acid
;; TITLE OF INVENTION: Molecules, and Uses Thereof
;; FILE REFERENCE: HW-8
;; CURRENT APPLICATION NUMBER: US/09/323,427
;; CURRENT FILING DATE: 1999-06-01
;; EARLIER APPLICATION NUMBER: 60/087,435
;; EARLIER FILING DATE: 1998-06-01
;; NUMBER OF SEQ ID NOS: 18
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO: 18
;; LENGTH: 892
;; TYPE: DNA
;; ORGANISM: Brugia malayi
US-09-323-427-18

Query Match 39.7%; Score 461; DB 4; Length 892;
Best Local Similarity 78.1%; Pred. No. 1,1e-124;
Matches 554; Conservative 0; Mismatches 155; Indels 0; Gaps 0;

QY 21 TTTCGTACTACACTTATGCAATGCTTATTCGATTCGCGTTGACCAATGCTGCAAG 80
DB 709 TTTCGTACTATGATATGCAAGATTTATGCTATTCGAATGATGCTGCAAG 650
QY 81 TGAGCAGAAATTTGATGTCGACCACTTCATTAACATCAATTTTAATACAGTATGC 140
DB 649 TGAACCTGAAATTTGATGTCGACCACTTCATTAATTTTAATACAGTATGC 590
QY 141 ATTCCAGAGACATGTTATGTAAGGCTTTATGATGTAAGAGTGGCGTAATGATGA 200
DB 589 TTTTGAAGACATGATATGTAAGGATTTATGATGTAAGATGTCGATGATGA 530
QY 201 AGGTGAGAGCTCAAGTGGCGGAATTTCACTTCATTTGATTCAGCAATGTCGCGTAC 260
DB 529 AGGTGAGAGCTCAAGTGGCGGAATTTCACTTCATTTGATTCAGCAATGTCGCGTAC 470
QY 261 AGCATCTGCAATCCAGTGGTATTTTGTAAACAACAACGCTGTCATTTGCTTCATCC 320
DB 469 AGCTTCGTTAAATCCAGTGGTATTTGTCACAGCTGTCGTAATTAATGCTTCATCC 410
QY 321 ATTATTTGTTACCAAGTTGATGTCGATGATGATGATGATGATGATGATGATGATGATGAT 380
DB 409 ACAGTTTATCAAAAAGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 540
QY 381 TAAACAGTTAGTGCACAGATTTGATGATGATGATGATGATGATGATGATGATGATGAT 440
DB 349 TAAGACTGTTAGCACACAATTTGAAGTTCCGAATTTGACAAACCGTATTTGCTACCAATT 290
QY 441 TGTCCTGATGCCAGTATCCGTTATGAATTTTGGATGTCGACCAACCGGTCACAGT 500
DB 289 GGTACCAATGCTGCTGTAGATATGATGATGATGATGATGATGATGATGATGATGATGAT 660
QY 501 TCAATTTGCTATCATTTGTCAGCCGTTATCATTAATGATGATGATGATGATGATGATGAT 560
DB 229 CCAATGCTATCATTTGTCAGCCGTTATCATTAATGATGATGATGATGATGATGATGATGAT 720
QY 561 TGATACCTTTCGCGGTTGTCATTCCTGCTTGTGATGATGATGATGATGATGATGATGAT 620
DB 169 TGATACCTTTCGCGGTTGTCATTCCTGCTTGTGATGATGATGATGATGATGATGATGAT 780
QY 621 GGAATTTCTAATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 680
DB 109 TAAATTTAATTAATGAAGAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
QY 681 TCCAAAGATTTAATGCTGTCGCAAGAAGCTCAGCTATACAAATATGCG 729
DB 49 TCCAAAGATTTAATGCTGTCGCAAGAAGCTCAGCTATACAAATATGCG 1

RESULT 15

US-09-812-642-16
; Sequence 16, Application US/09812642
; Patent No. 6368600
; GENERAL INFORMATION:
; APPLICANT: Chandrasekar, Ramaswamy
; APPLICANT: Morales, Tony H.
; TITLE OF INVENTION: Parasitic Helminth Cuticlin Proteins, Nucleic Acid
; FILE REFERENCE: Molecules, and Uses Thereof
; CURRENT APPLICATION NUMBER: US/09/812,642
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: 09/323,427
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 16
; LENGTH: 892
; TYPE: DNA
; ORGANISM: Brugia.malay1
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (158)..(892)
US-09-812-642-16

Query Match 39.7%; Score 461; DB 4; Length 892;
Best Local Similarity 78.1%; Pred. No.1.le-124;
Matches 554; Conservative 0; Mismatches 155; Indels 0; Gaps 0;

QY 21 TTTCTGACTACACTTATGCAATGCTTATGCAATCCGTTGCAATGGTGTGCAAG 80
DB 184 TTGTGATATATGATATATACAGATATATATGCTATTCATGATATGTTGCGAAG 243
QY 81 TGAGCCGAAATGAAATGTCGACCACTTCAATACAAATATTTAATACAGTATGC 140
DB 244 TGAACCTGAAATGAAATGTCGACCACTTCAATACAAATATTTAATACAGTATGC 303
QY 141 ATTGGAAGACATGTTATGTAAGGCTTTATGATCAAGAAGGTTCCGTAATGATGA 200
DB 304 TTTTGAAGGACATGTTATGTAAGGCTTTATGATCAAGAAGGTTCCGTAATGATGA 363
QY 201 AGGTGACGTCAGATGTCGCGGAATTTGCTCCATTTGATGATGCAATGTTGCGCTAC 260
DB 364 AGGTGACGTCAGATGTCGCGGAATTTGCTCCATTTGATGATGCAATGTTGCGCTAC 423
QY 261 AGCATCTGTAATCCAGTGTGTTTGTATGACAAACAGTGTGCTCTTTCGTTCAATCC 320
DB 424 ACCTTGTATTAATCCAGTGTGTTTGTATGACAAACAGTGTGCTCTTTCGTTCAATCC 483
QY 321 ATTATTTGTATCCAAAGTTGATGTCGATATGCAATGCTTTTACATGGAAGCTGA 380
DB 484 ACATTTTATCACAAAGTTGATGTCGATATGCAATGCTTTTACATGGAAGCTGA 543
QY 381 TAAACAGTTAGTCGACAGATGTAAGTATCTGAATATCACAATGCTTTTCAAACTCAAT 440
DB 544 TAAGACTGTTAGACACAAATTTGAAGTTCCGAATGACAACTATTTCTACACAAAT 603
QY 441 TGCCCGATCCAGTATGCGGTTATGAAATTTGGATGGGACCAACCGGTCAACAGT 500
DB 604 GGTACCAATGCTGTGTGTATGATATGATGATGATGATGATGATGATGATGATGAT 663
QY 501 TCAATTTGCTATATGTCAGCCAGTTTATCATTAATGAGCATGCAATGCTGTAACCGT 560
DB 664 CCAGTATGCTAATATGAGCAACCGGTTTATCATTAATGAGCATGCAATGCTGTAACCGT 723
QY 561 TGATCTTTTGTGCGGCTTTCATCTCTGCTTTGTCGATGATGATGATGATGATGAT 620
DB 724 TGATCTTTTGTGCGGCTTTCATCTCTGCTTTGTCGATGATGATGATGATGATGAT 783
QY 621 GGAATTTCTAATGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 680
DB 784 TAAATTTAATTAATGAGCAAGATGATGATGATGATGATGATGATGATGATGATGAT 843
QY 681 TCCAAAGATTTAATGCTGCGCAAGAAGCTCAGTATACAAATATGCG 729

DB 844 TCCAAAGATTTAATGCTGCGCAAGAAGCTCAGTATACAAATATGCG 892

RESULT 16
US-09-812-642-18/C
; Sequence 18, Application US/09812642
; Patent No. 6368600
; GENERAL INFORMATION:
; APPLICANT: Chandrasekar, Ramaswamy
; APPLICANT: Morales, Tony H.
; TITLE OF INVENTION: Parasitic Helminth Cuticlin Proteins, Nucleic Acid
; FILE REFERENCE: Molecules, and Uses Thereof
; CURRENT APPLICATION NUMBER: US/09/812,642
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: 09/323,427
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 18
; LENGTH: 892
; TYPE: DNA
; ORGANISM: Brugia.malay1
US-09-812-642-18

Query Match 39.7%; Score 461; DB 4; Length 892;
Best Local Similarity 78.1%; Pred. No.1.le-124;
Matches 554; Conservative 0; Mismatches 155; Indels 0; Gaps 0;

QY 21 TTTCTGACTACACTTATGCAATGCTTATGCAATCCGTTGCAATGGTGTGCAAG 80
DB 709 TTGTGATATATGATATATACAGATATATATGCTATTCATGATATGTTGCGAAG 650
QY 81 TGAGCCGAAATGAAATGTCGACCACTTCAATACAAATATTTAATACAGTATGC 140
DB 649 TGAACCTGAAATGAAATGTCGACCACTTCAATACAAATATTTAATACAGTATGC 590
QY 141 ATTGGAAGACATGTTATGTAAGGCTTTATGATCAAGAAGGTTCCGTAATGATGA 200
DB 589 TTTTGAAGGACATGTTATGTAAGGCTTTATGATCAAGAAGGTTCCGTAATGATGA 530
QY 201 AGGTGACGTCAGATGTCGCGGAATTTGCTCCATTTGATGATGCAATGTTGCGCTAC 260
DB 529 AGGTGACGTCAGATGTCGCGGAATTTGCTCCATTTGATGATGCAATGTTGCGCTAC 470
QY 261 AGCATCTGTAATCCAGTGTGTTTGTATGACAAACAGTGTGCTCTTTCGTTCAATCC 320
DB 469 ACCTTGTATTAATCCAGTGTGTTTGTATGACAAACAGTGTGCTCTTTCGTTCAATCC 410
QY 321 ATTATTTGTATCCAAAGTTGATGTCGATATGCAATGCTTTTACATGGAAGCTGA 380
DB 409 ACATTTTATCACAAAGTTGATGTCGATATGCAATGCTTTTACATGGAAGCTGA 350
QY 381 TAAACAGTTAGTCGACAGATGTAAGTATCTGAATATCACAATGCTTTTCAAACTCAAT 440
DB 349 TAAGACTGTTAGACACAAATTTGAAGTTCCGAATGACAACTATTTCTACACAAAT 290
QY 441 TGCCCGATCCAGTATGCGGTTATGAAATTTGGATGGGACCAACCGGTCAACAGT 500
DB 289 GGTACCAATGCTGTGTGTATGATATGATGATGATGATGATGATGATGATGATGAT 230
QY 501 TCAATTTGCTATATGTCAGCCAGTTTATCATTAATGAGCATGCAATGCTGTAACCGT 560
DB 229 CCAGTATGCTAATATGAGCAACCGGTTTATCATTAATGAGCATGCAATGCTGTAACCGT 170
QY 561 TGATCTTTTGTGCGGCTTTCATCTCTGCTTTGTCGATGATGATGATGATGATGAT 620
DB 169 TGATCTTTTGTGCGGCTTTCATCTCTGCTTTGTCGATGATGATGATGATGATGAT 110
QY 621 GGAATTTCTAATGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 680
DB 109 TAAATTTAATTAATGAGCAAGATGATGATGATGATGATGATGATGATGATGATGAT 50

Db 1486 CAAAAGATTAGACAGATATCCAGTTAAGTTAAAGCATACGAAGATGAACAGCTTCT 1545
QY 721 AATATCGGATCGATACAGCTTTCTATCAATGCCAGATGATATACATTAAGAA 780
Db 1546 ATTAAAGCTGCACGTGCGAAGACTTGAAAAACATATAAATGAAGACGAAACTTAACAGAA 1605
QY 781 CCA 783
Db 1606 CCA 1608

RESULT 24
US-09-323-427-11
; Sequence 11, Application US/09323427
; Patent No. 6248329
; GENERAL INFORMATION:
; APPLICANT: Chandrashekar, Ramaswamy
; TITLE OF INVENTION: Parasitic Helminth Cuticlin Proteins, Nucleic Acid
; FILE REFERENCE: HW-8
; CURRENT APPLICATION NUMBER: US/09/323,427
; EARLIER FILING DATE: 1999-06-01
; PRIOR APPLICATION NUMBER: 60/087,435
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 11
; LENGTH: 34
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-323-427-11

Query Match 2.9%; Score 34; DB 4; Length 34;
Best Local Similarity 100.0%; Pred. No. 0.47;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 696 GGCTGGCCAGAGCTCAGTATACAAATATGCG 729
Db 1 GGCTGGCCAGAGCTCAGTATACAAATATGCG 34

RESULT 25
US-09-323-427-12/C
; Sequence 12, Application US/09323427
; Patent No. 6248329
; GENERAL INFORMATION:
; APPLICANT: Chandrashekar, Ramaswamy
; APPLICANT: Morales, Tony H.
; TITLE OF INVENTION: Parasitic Helminth Cuticlin Proteins, Nucleic Acid
; FILE REFERENCE: HW-8
; CURRENT APPLICATION NUMBER: US/09/323,427
; CURRENT FILING DATE: 1999-06-01
; EARLIER APPLICATION NUMBER: 60/087,435
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 12
; LENGTH: 34
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-323-427-12

Query Match 2.9%; Score 34; DB 4; Length 34;
Best Local Similarity 100.0%; Pred. No. 0.47;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 696 GGCTGGCCAGAGCTCAGTATACAAATATGCG 729
Db 34 GGCTGGCCAGAGCTCAGTATACAAATATGCG 1

RESULT 26
US-09-812-642-11
; Sequence 11, Application US/09812642
; Patent No. 6368600
; GENERAL INFORMATION:
; APPLICANT: Chandrashekar, Ramaswamy
; APPLICANT: Morales, Tony H.
; TITLE OF INVENTION: Parasitic Helminth Cuticlin Proteins, Nucleic Acid
; FILE REFERENCE: HW-8
; CURRENT APPLICATION NUMBER: US/09/812,642
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: 09/323,427
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 11
; LENGTH: 34
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-812-642-11

Query Match 2.9%; Score 34; DB 4; Length 34;
Best Local Similarity 100.0%; Pred. No. 0.47;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 696 GGCTGGCCAGAGCTCAGTATACAAATATGCG 729
Db 34 GGCTGGCCAGAGCTCAGTATACAAATATGCG 1

RESULT 27
US-09-812-642-12/C
; Sequence 12, Application US/09812642
; Patent No. 6368600
; GENERAL INFORMATION:
; APPLICANT: Chandrashekar, Ramaswamy
; APPLICANT: Morales, Tony H.
; TITLE OF INVENTION: Parasitic Helminth Cuticlin Proteins, Nucleic Acid
; FILE REFERENCE: HW-8
; CURRENT APPLICATION NUMBER: US/09/812,642
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: 09/323,427
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 12
; LENGTH: 34
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-812-642-12

Query Match 2.9%; Score 34; DB 4; Length 34;
Best Local Similarity 100.0%; Pred. No. 0.47;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;


```
;
;   REGISTRATION NUMBER: 29,768
;   REFERENCE/DOCKET NUMBER: 16754/115 CHCO
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (703)836-9300
;   TELEFAX: (703)683-4109
;   TELEX: 899149
;   INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1867 base pairs
;   TYPE: NUCLEIC ACID
;   STRANDEDNESS: double
;   TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
;   ORIGINAL SOURCE:
;   STRAIN: 164A-1
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 716..1849
;   OTHER INFORMATION: /product= "mature protein of
;   OTHER INFORMATION: 164A-1"
;   FEATURE:
;   NAME/KEY: sig-peptide
;   LOCATION: 716..1021
;   FEATURE:
;   NAME/KEY: mat-peptide
;   LOCATION: 1022..1846
;   US-07-772-087-1
```

```
Query Match          2.9%; Score 33.4; DB 1; Length 1867;
Best Local Similarity 52.5%; Pred. No. 5.4;
Matches 96; Conservative 0; Mismatches 86; Indels 1; Gaps 1;
```

```
QY 601 GATGCTACGCTGATCTGCGGAATCTTAATCTGATGATGCTCTTGATTAATAT 660
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 794 GAAGCAAGTGGAAATGCTGAATTAAGAAAGATTACTTAGTTGTTTAAACGAAAT 853
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 661 TTGCTAATTAATTTGAATATCCACAGATTTAATGCTGGCCAGAAGCTCAGCTATAC 720
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 854 GTAACCATGATCTAATAAATATATAGTAACCTCACTAGTGAAGTGTACATCAATAT 913
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 721 AATATGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 780
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 914 AAGTATATGACGATATGCA-TGTATCTCTCCAGAAAAAGACAGACAGATTAATAAAA 972
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 781 CCA 783
      |||
DB 973 CAA 975
```

RESULT 32

```
US-09-453-702B-166
; Sequence 166, Application US/09453702B
; Patent No. 6365723
; GENERAL INFORMATION:
;   APPLICANT: Blatner, Frederick R.
;   Burland, Valerie
;   Perna, Nicole T.
;   Plunkett, Guy
;   Welch, Rod
;   TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157
;   NUMBER OF SEQUENCES: 265
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Quarles & Brady
;   STREET: 1 South Pinckney Street
;   CITY: Madison
;   STATE: WI
;   COUNTRY: US
;   ZIP: 53701-2113
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette, 3.50 inch, 1.44MB storage
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: Word Perfect 8.0
```

```
;
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/453,702B
;   FILING DATE: 03-Dec-1999
;   CLASSIFICATION: <Unknown>
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 60/110,955
;   FILING DATE: 04-DEC-1998
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Seay, Nicholas J.
;   REGISTRATION NUMBER: 27386
;   REFERENCE/DOCKET NUMBER: 960296.95017
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (608) 251-5000
;   TELEFAX: (608) 251-9166
;   INFORMATION FOR SEQ ID NO: 166:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 16950
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
;   SEQUENCE DESCRIPTION: SEQ ID NO: 166:
;   US-09-453-702B-166
```

```
Query Match          2.9%; Score 33.4; DB 4; Length 16950;
Best Local Similarity 51.7%; Pred. No. 17;
Matches 76; Conservative 0; Mismatches 71; Indels 0; Gaps 0;
```

```
QY 405 GGTATCTGAATTCACCACTGCTTTTCAACATTCGAATGTCGCCGATGCGCGTTA 464
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1417 GGTGCTTAATTCCTCCACAGCTTTTTCACACTGCTGCGTTAGATCAATCAACGCCCTCGA 1476
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 465 TGAATTTTGGATGCTGACCAACCGGTCACACAGTTCAATTTGCTATCATTTGTCAGCC 524
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1477 TCAATTTTGGTCTGATGCTCCATATATTAACCATTCATCTGCATAGTCCCTCCTCA 1536
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 525 AGTTATCATTAATGACATGCGATTC 551
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1537 ATTATTCCTGAATATGTCACCTGCTAC 1563
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 33

```
5182210-9/c
; Patent No. 5182210
; APPLICANT: BINNS, MATTHEW M.;BOURNSNELL, MICHAEL E.G.;
; CAMPBELL, JOAN I.A.;TOMLEY, FIONA M.
; TITLE OF INVENTION: FOWLPOX VIRUS PROMOTERS
; NUMBER OF SEQUENCES: 22
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/469,608
;   FILING DATE: 21-OCT-1988
;   SEQ ID NO:9:
;   LENGTH: 11225
;   5182210-9
```

```
Query Match          2.8%; Score 33; DB 6; Length 11225;
Best Local Similarity 55.6%; Pred. No. 18;
Matches 85; Conservative 0; Mismatches 65; Indels 3; Gaps 1;
```

```
QY 290 TAAACAACATGTTGATATTTGTTTCATCATTAATTTGTTACCAAGTGTGTCAT 349
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 9172 TAGAACAATCTTATTAACAGTAATTAATGACTTATTTACTTAACAAAGCTGAAGAGATAT 9113
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 350 ATCGAGTACAAATGCTTTTACATGGAAGCTGATAAACAAGTTAGTGCACAGATTGAGTAT 409
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 9112 ---TAGAAAAAGACTATTTCTATGGAACAGATATATGAATGAAGAGATATCCGTACAG 9056
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 410 CTGAATCACACATGCTTTTCAAACTCAATTTG 442
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 9055 ATGAATGATAGCTGCTATAGAGCTGAATTTG 9023
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 34

Db 186 TCCGTAATAAATGCAATTTCCACCGCCTTTTGGCTTGACATAGCATGGGATCAG 127

QY 1098 CATGCATTCATTCCTGCGCATTTATTCATTCCTGTTAAATTT 1143

Db 126 CATCATCGGGCGGAATAGGAAAAAATATCGCTCGATAAACTT 81

RESULT 37

US-08-781-802-5/c

; Sequence 5, Application US/08781802

; Patent No. 5969121

; GENERAL INFORMATION:

; APPLICANT: ALLEN, Larry

; APPLICANT: AIKENS, John

; APPLICANT: FONSTEIN, Michael

; APPLICANT: VONSTEIN, Veronika

; APPLICANT: DEMIRJIAN, David

; APPLICANT: CASADABAN, Malcolm

; TITLE OF INVENTION: Stable Biocatalysts for Ester Hydrolysis

; NUMBER OF SEQUENCES: 12

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff

; STREET: 300 S. Wacker Drive 32nd Floor

; CITY: Chicago

; STATE: Illinois

; COUNTRY: USA

; ZIP: 60606

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/781,802

; FILING DATE: 10-JAN-1997

; CLASSIFICATION: 536

; APPLICATION NUMBER: US 08/694,078

; FILING DATE: 07-AUG-1996

; CLASSIFICATION: 536

; APPLICATION NUMBER: US 60/019,580

; FILING DATE: 12-JUN-1996

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 60/009,704

; FILING DATE: 11-JAN-1996

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 60/001,995

; FILING DATE: 01-AUG-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Chao, Mark

; REGISTRATION NUMBER: 37,293

; REFERENCE/DOCKET NUMBER: 95,963-E

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 312-913-0001

; TELEFAX: 312-913-0002

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 4090 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 197..1699

; OTHER INFORMATION: /note="E011 sequence of longest

; OTHER INFORMATION: open reading frame; other possible start codons ATG/met4;

; OTHER INFORMATION: TTG/leu7; GTG/val8; GTG/val15; GTG/val36; ATG/met52"

; NAME/KEY: mat-peptide

; LOCATION: 197..1699

US-08-781-802-5

Query Match

Best Local Similarity 2.8%; Score 32.4; DB 2; Length 4090;

Matches 129; Conservative 0; Mismatches 161; Indels 0; Gaps 0;

QY 150 ACATGTTATGTAAGAGTCTTTATGATCAGAAAGTGGCCGTATGATGAAGTGGACG 209

Db 2895 ACATGTTATGTAAGAGTCTTTATGATCAGAAAGTGGCCGTATGATGAAGTGGACG 2836

QY 210 TCAAGTTCGCGAATTCCTTCATTCCTTCATTCATGATGTCGCGCTACACATCTCT 269

Db 2835 ACAAGTTACTTGGTGAACACGACATTTGACTCTCAGACATGAAAAATGTGGGAA 2776

QY 270 GAATTCACGCTGATTTTGTGAACACACACTGTGTCTATTTGTTTCATCCATTTTGT 329

Db 2775 ATATCCGATCCGCAACCCCTTTGTAATAAAGGGGTTCCGCAATTTATCTTTATATAGT 2716

QY 330 TACCAAGTTGATGTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 389

Db 2715 TACAAAGAAATGTTTGTGTAACCTGTAAGGCGCAATAATGAGTGGATGAGAGGAGG 2656

QY 390 TAGTGCACAGATTGAGTATCTGAATCACAACCTCTTTCAAACTCAA 439

Db 2655 AGAAGGAACCCGCGCTGTTTCGCGCCACAATATTTCTTTATATCAGA 2606

RESULT 38

US-08-694-078-5/c

; Sequence 5, Application US/08694078

; Patent No. 6218163

; GENERAL INFORMATION:

; APPLICANT: ALLEN, Larry

; APPLICANT: AIKENS, John

; APPLICANT: FONSTEIN, Michael

; APPLICANT: VONSTEIN, Veronika

; APPLICANT: DEMIRJIAN, David

; APPLICANT: CASADABAN, Malcolm

; TITLE OF INVENTION: Stable Biocatalysts for Ester Hydrolysis

; NUMBER OF SEQUENCES: 8

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff, Ltd.

; STREET: 300 S. Wacker Drive 7th Floor

; CITY: Chicago

; STATE: Illinois

; COUNTRY: USA

; ZIP: 60606

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/694,078

; FILING DATE: 07-AUG-1996

; CLASSIFICATION: 435

; APPLICATION NUMBER: US 60/019,580

; FILING DATE: 12-JUN-1996

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 60/009,704

; FILING DATE: 10-JAN-1996

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 60/001,995

; FILING DATE: 07-AUG-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Chao, Mark

; REGISTRATION NUMBER: 37,293

; REFERENCE/DOCKET NUMBER: 95,963-C

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 312-913-0001

; TELEFAX: 312-913-0002

; INFORMATION FOR SEQ ID NO: 5:

Db 268 ACAAGTTGGGCGACACATGGTGTCCAAATTATGAAATTCACATCCACACACAGTCACAA 327

OY 1024 CCTGTAACTCTGCTGCAGTACAAAATGG 1052

Db 328 TCTGACGTTTACATTAAGTTCAATAATGG 356

Search completed: March 28, 2003, 17:00:55
Job time : 141 secs

THIS PAGE BLANK (USPTO)


```

; Sequence 5 Application US/10054562A
; Patent No. US20020165375A1
; GENERAL INFORMATION:
; APPLICANT: Chandrashekar, Ramaswamy
; APPLICANT: Morales, Tony H.
; TITLE OF INVENTION: PARASITIC HELMINTH CUTICLIN PROTEINS, NUCLEIC ACID MOLECULES,
; FILE REFERENCE: HW-8-2
; CURRENT APPLICATION NUMBER: US/10/054,562A
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: 09/812,642
; PRIOR FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: 09/323,427
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 1161
; TYPE: DNA
; ORGANISM: Dirofilaria immitis
US-10-054-562A-5

Query Match      100.0%; Score 1161; DB 9; Length 1161;
Best Local Similarity 100.0%; Pred. No. 4.7e-294;
Matches 1161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY   1 ATATGATTCTGGTTTATGTTTCGTACTACACCTATTGCATTTGCTATTGCGANTCCG 60
Db   1161 AATATCATTCGTCCTTAATTCCTCCTCTACTACTACTATTGTCATTGCTATTGCGATCCG 1102

QY   61 GTTGACAATGCTGTGCAAGGTGACGCCAGAATTGAATGTGACCACCTTCATAAACAAATC 120
Db   1101 GTTGACAAATGCTGTGCAAGGTGACGCCAGAAATGAATGTGACCACCTTCATAAACAAATC 1042

QY   121 AATTTAATAACGCTAATCATTCGAGAGCAATGTTTAGTGAAGGCTTTATGATCAA 180
Db   1041 AATTTAATAACGCTAATCATTCGAGAGCAATGTTTAGTGAAGGCTTTATGATCAA 982

QY   181 GAAGTTGGCCGTAATGATGAAGGTGACGTCGAAGTTGCCGGAATTTCACTTCATTGAT 240
Db   981 GAAGTTGGCCGTAATGATGAAGGTGACGTCGAAGTTGCCGGAATTTCACTTCATTGAT 922

QY   241 TCATGAAATGTGGCGGTACAGATCTCTGAATCCAAGTGATTTTGTGAACAACAAT 300
Db   921 TCATGAAATGTGGCGGTACAGATCTCTGAATCCAAGTGATTTTGTGAACAACAAT 862

QY   301 GTTGTCATTTCTTTCATCATCATTTATTTGTACCAAGTTGATCGGCAATGAGATACAA 360
Db   861 GTTGTCATTTCTTTCATCATCATTTATTTGTACCAAGTTGATCGGCAATGAGATACAA 802

QY   361 TGCCTTTACATGGAAAGCTGATAAAACAGTTAGTGACAGATTTAGAGTATCTGAATACCA 420
Db   801 TGCCTTTACATGGAAAGCTGATAAAACAGTTAGTGACAGATTTAGAGTATCTGAATACCA 742

QY   421 ACTGCTTTCAACGCAATTTGTCCCGAATGCCAGATGCCGTTATGAAATTTTGGATGCT 480
Db   741 ACTGCTTTCAACGCAATTTGTCCCGAATGCCAGATGCCGTTATGAAATTTTGGATGCT 682

QY   481 GGACCAACCGGTCAACCAAGTTCAATTTGCTATCATTTGTGTCAACCGCAAGTTATCATTAATGG 540
Db   681 GGACCAACCGGTCAACCAAGTTCAATTTGCTATCATTTGTGTGTCAACCGCAAGTTATCATTAATGG 622

QY   541 ACATCGAATTCGAAACCGTTGATACCTTGTGCGCGGTTGTCATTTCCCTGCTTGTGCGAT 600
Db   621 ACATCGAATTCGAAACCGTTGATACCTTGTGCGCGGTTGTCATTTCCCTGCTTGTGCGAT 562

QY   601 GATGTTACCGGTGATCTGTGTAATTTCTAATGTGTATGATGTGCTCTTGATTAATAT 660
Db   561 GATGTTACCGGTGATCTGTGTAATTTCTAATGTGTATGATGTGCTCTTGATTAATAT 502

QY   661 TTGCTAAATTAATTTGGAATATCCAAAGATTAATTAAGTGTGGCCAAAGACACAGTATAC 720
Db   501 TTGCTAAATTAATTTGGAATATTCCAAAGATTAATTAAGTGTGGCCAAAGACACAGTATAC 442

```


QY 721 AATATGCGATGCATCAGCTTTTCTATCAATGCGAGATCAGTATTACCATTAAGAA 780
| | | | |
Db 441 AATATGCGATGCATCAGCTTTTCTATCAATGCGAGATCAGTATTACCATTAAGAA 382
| | | | |
QY 781 CCAATATGCGATGCTTGCAGCACAATGTTGAGAACACAGAGATTGGAGCTGTTAAA 840
| | | | |
Db 381 CCAATATGCGAATGTTGTCAGCACAATGTTGAGAACACAGAGATTGGAGCTGTTAAA 322
| | | | |
QY 841 ACAGGTGTGCGCGAGCAAAACCTGCTGCAGCTGCGCAACTGCTTTACTCAGAAAAGA 900
| | | | |
Db 321 ACAGGTGTGCGCGAGCAAAACCTGCTGCAGCTGCGCAACTGCTTTACTCAGAAAAGA 262
| | | | |
QY 901 TCGCAGAACCGGAGAAATATCATTTGATGTACGAACATGATCAACACCTTGAATATAGC 960
| | | | |
Db 261 TCGCAGAACCGGAGAAATATCATTTGATGTACGAACATGATCAACACCTTGAATATAGC 202
| | | | |
QY 961 GATGATTAATCAAGCTTTGCCAGTTGATTTAGCTCACCGTGCACCTTCTGCAACATTAATGA 1020
| | | | |
Db 201 GATGATTAATCAAGCTTTGCCAGTTGATTTAGCTCACCGTGCACCTTCTGCAACATTAATGA 142
| | | | |
QY 1021 CAACCTGTATATCTGCTGCAGTACAAATGGAATCTGCATCTGCACATTTGGCTTCTCA 1080
| | | | |
Db 141 CAACCTGTATATCTGCTGCAGTACAAATGGAATCTGCATCTGCACATTTGGCTTCTCA 82
| | | | |
QY 1081 ATGTTATAGGTTTAAAGCATTTGATGATGCTGCGCGCTCATTTATACCATTTGGTTAAA 1140
| | | | |
Db 81 ATGTTATAGGTTTAAAGCATTTGATGATGCTGCGCGCTCATTTATACCATTTGGTTAAA 22
| | | | |
QY 1141 TTTCGTCCAATCAGAAGGCA 1161
| | | | |
Db 21 TTTCGTCCAATCAGAAGGCA 1

RESULT 3
US-10-054-562a-1
; Sequence 1, Application US/10054562a
; Patent No. US20020165375a1
; GENERAL INFORMATION:
; APPLICANT: Chandrasekar, Ramaswamy
; TITLE OF INVENTION: PARASITIC HELMINTH CUTICLIN PROTEINS, NUCLEIC ACID MOLECULES, ANT
; FILE REFERENCE: HW-8-2
; CURRENT APPLICATION NUMBER: US/10/054,562a
; PRIOR FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: 09/812,642
; PRIOR FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: 09/323,427
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1779
; TYPE: DNA
; ORGANISM: Dicrofilaria immitis
US-10-054-562a-1

Query Match 100.0%; Score 1161; DB 9; Length 1779;
Best Local Similarity 100.0%; Pred. No. 6e-294;
Matches 1161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGATGATGCTGCTTATGCTTCTGTAACACTTATTCGATGCTTATTCGATTCGCG 60
| | | | |
Db 167 ATGATGATGCTGCTTATGCTTCTGTAACACTTATTCGATGCTTATTCGATTCGCG 226
| | | | |
QY 61 GTTGACAATGCTGTGCAAGGTGAGCCAGAAATGGAATGTGACCAACTTCAATTAACATC 120
| | | | |
Db 227 GTTGACAATGCTGTGCAAGGTGAGCCAGAAATGGAATGTGACCAACTTCAATTAACATC 286
| | | | |
QY 121 AATTTTAATACAGTAATGCAATTCGAAGAAGACATGTTATGTGAAGGCTTTTATGATCAA 180
| | | | |
Db 287 AATTTTAATACAGTAATGCAATTCGAAGAAGACATGTTATGTGAAGGCTTTTATGATCAA 346
| | | | |

QY 181 GAAAGTTGCCGTAATGATGAAGGTGAGCTCAAGTGTCCGGAATTTTCATCTCCATTGAT 240
| | | | |
Db 347 GAAAGTTGCCGTAATGATGAAGGTGAGCTCAAGTGTCCGGAATTTTCATCTCCATTGAT 406
| | | | |
QY 241 TCATGCAATGTGGCGCTACACAGATCTGTAATCCAGCTGTATTTTGTAAACAACAT 300
| | | | |
Db 407 TCATGCAATGTGGCGCTACACAGATCTGTAATCCAGCTGTATTTTGTAAACAACAT 466
| | | | |
QY 301 GTTGCTATTTGCTTTCATCTATTTTGTATACCAAAATGTGATGTGATTCAGATCAA 360
| | | | |
Db 467 GTTGCTATTTGCTTTCATCTATTTTGTATACCAAAATGTGATGTGATTCAGATCAA 526
| | | | |
QY 361 TCGCTTTACATGGAAGGTGATTAACAGTATAGTGCACAGATTTGAGTATCGAATATACA 420
| | | | |
Db 527 TCGCTTTACATGGAAGGTGATTAACAGTATAGTGCACAGATTTGAGTATCGAATATACA 586
| | | | |
QY 421 ACTGCTTTTGAACACTCAAAATTTGTCGATGCGCAGATATCCGTTATGAATTTTGGATGGT 480
| | | | |
Db 587 ACTGCTTTTGAACACTCAAAATTTGTCGATGCGCAGATATCCGTTATGAATTTTGGATGGT 646
| | | | |
QY 481 GGACCAACCGGTACACAGTTCATTTGCTATCATTTGTCACGCAATTTATCATTAATGC 540
| | | | |
Db 647 GGACCAACCGGTACACAGTTCATTTGCTATCATTTGTCACGCAATTTATCATTAATGC 706
| | | | |
QY 541 ACATGCGATTTGGAACCGTTGATCTTTGCGCGGTTGCGCATTCCTGCTTTGTGAT 600
| | | | |
Db 707 ACATGCGATTTGGAACCGTTGATCTTTGCGCGGTTGCGCATTCCTGCTTTGTGAT 766
| | | | |
QY 601 GATGCTAACGCTGATCTGTGGAATTTCTAATGCTGATGATGCTGCTTGATTAATAT 660
| | | | |
Db 767 GATGCTAACGCTGATCTGTGGAATTTCTAATGCTGATGATGCTGCTTGATTAATAT 826
| | | | |
QY 661 TTGCTTAATTAATTTGGAATTTCCACAGATTTAATGCTGCGCAAGAGCTCACGTTATC 720
| | | | |
Db 827 TTGCTTAATTAATTTGGAATTTCCACAGATTTAATGCTGCGCAAGAGCTCACGTTATC 886
| | | | |
QY 721 AATATGCGAATGATGATACAGCTTTTCTATCAATGCCAGATTCATTTACCATTAAGAA 780
| | | | |
Db 887 AATATGCGAATGATGATACAGCTTTTCTATCAATGCCAGATTCATTTACCATTAAGAA 946
| | | | |
QY 781 CCAATATGCGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
| | | | |
Db 947 CCAATATGCGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1006
| | | | |
QY 841 ACAGGTGTGCGCGAGCAAAACCTGCTGCAGCTGCGCAACTTCTTACTCAGAAAAGA 900
| | | | |
Db 1007 ACAGGTGTGCGCGAGCAAAACCTGCTGCAGCTGCGCAACTTCTTACTCAGAAAAGA 1066
| | | | |
QY 901 TCGCAGAACCGGAGTAATCATTTGATGATGATGATGATGATGATGATGATGATGATGAT 960
| | | | |
Db 1067 TCGCAGAACCGGAGTAATCATTTGATGATGATGATGATGATGATGATGATGATGATGAT 1126
| | | | |
QY 961 GATGATTAATCAAGCTTTGCCAGTTGATTTAGCTCACCGTGCACCTTCTGCAACATTAATGA 1020
| | | | |
Db 1127 GATGATTAATCAAGCTTTGCCAGTTGATTTAGCTCACCGTGCACCTTCTGCAACATTAATGA 1186
| | | | |
QY 1021 CAACCTGTATATCTGCTGCAGTACAAATGGAATCTGCATCTGCACATTTGGCTTCTCA 1080
| | | | |
Db 1187 CAACCTGTATATCTGCTGCAGTACAAATGGAATCTGCATCTGCACATTTGGCTTCTCA 1246
| | | | |
QY 1081 ATGTTATAGGTTTAAAGCATTTGATGATGCTGCGCGCTCATTTATACCATTTGGTTAAA 1140
| | | | |
Db 1247 ATGTTATAGGTTTAAAGCATTTGATGATGCTGCGCGCTCATTTATACCATTTGGTTAAA 1306
| | | | |
QY 1141 TTTCGTCCAATCAGAAGGCA 1161
| | | | |
Db 1307 TTTCGTCCAATCAGAAGGCA 1327
| | | | |

RESULT 4
US-10-054-562a-2/c
; Sequence 2, Application US/10054562a

Query Match	43.4%	Score 504.2	DB 9	Length 1372
Best Local Similarity	76.9%	Pred. No.5.1e-128		
Matches 652	Conservative	0	Mismatches 193	Indels 3
			Gaps	3
QY 1	ATGATGATGCTTATTCCTTATGCTTCTTGACATCACTTATG-CATTGCTTATTCAGATTC	59		
Db 1341	ATTATCATTTATTCCTCCTTGTTATTCCTACTTCTTACTTGACCTACCTAGTATCATGATCC	1282		
QY 60	GCTTGACAAATGCTGTCGAAGGTGAGCCAGAAATGGAATGTGACCAATTCATTAACAT	119		
Db 1281	TATTGACAAATGCTGTCGAAGGTGAGCCAGAAATGGAATGTGAGCGCAGCTTCGATTAACAT	1222		
QY 120	CAATTTTAAATACAGTAATGATTCGAGACATGTTATGTGAAAGGCTTATATATCA	179		
Db 1221	CAATTTTAAATACAGTAATGATTCGAGACATGTTATGTGAAAGGCTTATATATCA	1162		
QY 180	AGAAGTGTGCGGTATATGAAAGTGAAGTGAAGTGCAGGTAATTCATCTTCATTTGA	239		
Db 1161	GGATGATATGCTTCACATATAGTATATGACCGGACAGTACGTAATGGAATGGAATGGA	1102		
QY 240	TTTCATGCAATGCTGCGGTACACGATCTGGAATCCACGCTGTAATTTTGTAAACAAC	299		
Db 1101	TTTCGTAAATGTTGAAGATACAGATCCCTTAATCCCTCGTGCTTTTGTAAACAACGT	1042		
QY 300	TGTTGCTATTTCTTTATCCATTCATTTATTTGTACCAAGTGAATCGTGCATATGATATCA	359		
Db 1041	AGTTGTATTTATTCATTTATCCCAAAATTCGTTAAAAAATGATGAGCAATTCGTATGA	982		
QY 360	ATGCTTTTACATG-AAGCTGATTAACACAGTTTATGTC-ACAGATATAGATTCGAATC	417		
Db 981	ATGTTTTTATATGAAAGCTGATATAGACCTTATGTAAGTCTTGAAGTATCGAAATG	922		
QY 418	ACAACCTCTTTTCAACTCAAAATTTGCCGATGCCAGATGCCGTTATGAAATTTTGAT	477		
Db 921	ACTACACATTCACAACTCAAGTGTACCAATGCCGATGTCGATATGATGATGATTTGGAA	862		
QY 478	GGTGACCACACCGGTCAACAGTTCAATTTGATATATTTGTCAGCCAGTTTATCAATAA	537		
Db 861	GGTGACCACACCGGTGCAACCTGTCGATTTGCAATGATTTGGAGATATGATATACAA	802		
QY 538	TGACATGACGATTTTGAACCGTTGATACATTTCTGCGCGGTTTCCATTCCTGCTTGTG	597		
Db 801	TGGACATGTTGATTCAGAGACTACAGATACATTCCTGTCATTAATCATTTGATGTTGG	742		
QY 598	GATATGCTAACGCTGATCTGTCGAAATTCCTAAATGCTGATGATGCTCTTGATAA	657		
Db 741	GATATGGAAGAAAGGTGATGTCAGTGCATGATTCCTAATGAAAGAGTGTGCTTTGGACAA	682		
QY 658	TATTTGTAATATTTTGAATATCCAAAGATTTTATGCTGCGCCAGAAAGCTCAGTA	717		
Db 681	TATTTTACATTAATTTTGAATATATTCAGATTTTATGCTGCGCCAGAAAGCTCAGTT	622		
QY 718	TACAAATATGCGGATGATCATCAGCTTTTATCAATGCCAGATTCAGTATTAACATTA	777		
Db 621	TATTAATATGAGATTCGATCAGAACTTTACTATCAATGCCAGATTTGATTAACAATTA	562		
QY 778	GAACCAATATGCGATGTTGTCGACCAATTTTCAGACACCAAGAGATTGCGAGCTGT	837		
Db 561	GAGCACAATATGCGATGCTTCGACCAATTCGACAGAGCCACAAGAGATTGCTGCATA	502		
QY 838	AAAACAGS 845			
Db 501	AAATCTGG 494			

```
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: HW-8-2
; CURRENT APPLICATION NUMBER: US/10/054,562A
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: 09/812,642
; PRIOR FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: 09/323,427
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 16
; LENGTH: 892
; TYPE: DNA
; ORGANISM: Brugia malayi
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (158)..(892)
; OTHER INFORMATION:
US-10-054-562A-16
```

```
Query Match          39.7%; Score 461; DB 9; Length 892;
Best Local Similarity 78.1%; Pred. No. 8.2e-111;
Matches 554; Conservative 0; Mismatches 155; Indels 0; Gaps 0;
```

```
QY 21 TTCTGTACTACACTATTCGCTTATTCGATCCGTTGACAAATGCTGCAAG 80
DB 184 TTGTGATATATGATATATACAGATATTAATGCTATTCGAATGAAATGGTCCAAAG 243
QY 81 TGAGCCAGAAATGGAATGAGGACCACTTCAATTAACATCAATTTTAATACAGTATGC 140
DB 244 TGAACCTGAAATGGAATGAGGACCACTTCAATTAACATCAATTTTAATACAGTATGC 303
QY 141 ATTGAGAGCAATGTTTATGTAAGAGCTTTATGATCAAGAAGGTGCGTAATGATGA 200
DB 304 TTTTGAAGGACATGATATGCTAAAGATATACAGTAATCAAGATTTGCTAGTATGA 363
QY 201 AGGTGACGCTCAAGTGGCCGAATTTCACTTCATTTGATCAATGAAATGTCGGCTAC 260
DB 364 AGGTGACGCTCAAGTGGCCGAATTTCACTTCATTTGATCAATGAAATGTCGGCTAC 423
QY 261 AGCATCTGGAATCCACAGTGTATTTTGTAAACAACAAGTGTGCTTTCATTCGTTATCC 320
DB 424 ACCTTGTAATATCCACAGTGTATTTTGTAAACAACAAGTGTGCTTTCATTCGTTATCC 483
QY 321 ATTATTTGTACCAAGTTGATCGCATATGAGTACATGCTTTTACATGGAAGCTGA 380
DB 484 ACAGTTTATCAAAAAGTTGATCGCATATGAGTACATGCTTTTACATGGAAGCTGA 543
QY 381 TAAACAGTTAGTGCACAGATGAGTATCTGAAATCAACAAGTGTTCCTTCAAACTCAAT 440
DB 544 TAAAGCTGTATGACACCAAAATGGAAGTTCGAAATGACACAGTATTTGCTACACAAT 603
QY 441 TGTCCGATGCCAGTATGCGCTTATGAAATTTTGTGATGAGGACCAACCGGTCAACAGT 500
DB 604 GGTACCAATGCGTGTGTATGATGAGATTTCTGATGATGCGTCCACCGGACCAACCTGT 663
QY 501 TCAATTTGCTATCATTTGCTACAGCACTTTATCATTAATGACATGCGATGGAACCGT 560
DB 664 CCAGATGCTATATATGGAACAACCGTTTATCATTAATGACATGCGATGGAACAGT 723
QY 561 TGATCTTCTGCGCGGTTGCTCATTTCCGTTTGTGATGATGGAAGGAGTACAGT 620
DB 724 TGATCTTCTGCGCGGTTGCTCATTTCCGTTTGTGATGATGGAAGGAGTACAGT 783
QY 621 GGAATTTCAATGCTGATGATGCTCTTGTGATTAATTTGCTAAATTAATTTGGAATA 680
DB 784 TAAATTAATTAATGAAGAAGATGATGATTAATGATGATTAATTTGGAATA 843
QY 681 TCCACACATTTAATGCTGCGCAAGAAGCTACATGATTAATTAATTTGCG 729
DB 844 TCCACATGATTAATGCTGCGCAAGAAGCTACATGATTAATTAATTTGCG 892
```

```
RESULT 8
US-10-054-562a-18/c
; Sequence 18, Application US/10054562A
; Patent No. US20020165375A1
; GENERAL INFORMATION:
; APPLICANT: Chandrashekar, Ramaswamy
; APPLICANT: Morales, Tony H.
; TITLE OF INVENTION: PARASITIC HELMINTH CUTICLIN PROTEINS, NUCLEIC ACID MOLECULES,
; FILE REFERENCE: THEREOF
; CURRENT APPLICATION NUMBER: US/10/054,562A
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: 09/812,642
; PRIOR FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: 09/323,427
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 18
; LENGTH: 892
; TYPE: DNA
; ORGANISM: Brugia malayi
US-10-054-562A-18
```

```
Query Match          39.7%; Score 461; DB 9; Length 892;
Best Local Similarity 78.1%; Pred. No. 8.2e-111;
Matches 554; Conservative 0; Mismatches 155; Indels 0; Gaps 0;
```

```
QY 21 TTCTGTACTACACTATTCGCTTATTCGATCCGTTGACAAATGCTGCAAG 80
DB 709 TTGTGATATATGATATATGCAAGATTAATGCTATTCGAATGAAATGGTCCAAAG 650
QY 81 TGAGCCAGAAATGGAATGAGGACCACTTCAATTAACATCAATTTTAATACAGTATGC 140
DB 649 TGAACCTGAAATGGAATGAGGACCACTTCAATTAACATCAATTTTAATACAGTATGC 590
QY 141 ATTGAGAGCAATGTTTATGTAAGAGCTTTATGATCAAGAAGGTGCGTAATGATGA 200
DB 589 TTTTGAAGGACATGATATGCTAAAGATTAACAGTAATCAAGATTTGCTAGTATGA 530
QY 201 AGGTGACGCTCAAGTGGCCGAATTTCACTTCATTTGATCAATGAAATGTCGGCTAC 260
DB 529 AGGTGACGCTCAAGTGGCCGAATTTCACTTCATTTGATCAATGAAATGTCGGCTAC 470
QY 261 AGCATCTGGAATCCACAGTGTATTTTGTAAACAACAAGTGTGCTTTCATTCGTTATCC 320
DB 469 ACCTTGTAATATCCACAGTGTATTTTGTAAACAACAAGTGTGCTTTCATTCGTTATCC 410
QY 321 ATTATTTGTACCAAGTTGATCGCATATGAGTACATGCTTTTACATGGAAGCTGA 380
DB 409 ACAGTTTATCAAAAAGTTGATCGCATATGAGTACATGCTTTTACATGGAAGCTGA 350
QY 381 TAAACAGTTAGTGCACAGATGAGTATCTGAAATCAACAAGTGTTCCTTCAAACTCAAT 440
DB 349 TAAAGCTGTATGACACCAAAATGGAAGTTCGAAATGACACAGTATTTGCTACACAAT 290
QY 441 TGTCCGATGCCAGTATGCGCTTATGAAATTTTGTGATGAGGACCAACCGGTCAACAGT 500
DB 289 GGTACCAATGCGTGTGTATGATGAGATTTCTGATGATGCGTCCACCGGACCAACCTGT 230
QY 501 TCAATTTGCTATCATTTGCTACAGCACTTTATCATTAATGACATGCGATGGAACCGT 560
DB 229 CCAGATGCTATATATGGAACAACCGTTTATCATTAATGACATGCGATGGAACAGT 170
QY 561 TGATCTTCTGCGCGGTTGCTCATTTCCGTTTGTGATGATGGAAGGAGTACAGT 620
DB 169 TGATCTTCTGCGCGGTTGCTCATTTCCGTTTGTGATGATGGAAGGAGTACAGT 110
QY 621 GGAATTTCAATGCTGATGATGCTCTTGTGATTAATTTGCTAAATTAATTTGGAATA 680
DB 109 TAAATTAATTAATGAAGAAGATGATGATTAATGATGATTAATTTGGAATA 50
QY 681 TCCACACATTTAATGCTGCGCAAGAAGCTACATGATTAATTAATTTGCG 729
```


Db 260 GATTGATCAAGACGATCAATTAATCTGATTAATAGGAGTGGAGTCCAGTCCGTTTAA 201
QY 944 ACACCTTGAATAGGATGATTAATCAAGCTTGGCCAGTTGAT 987
Db 200 GCACGGTTGATATTAACGAGAGATCCGAACCTTCTCAGCAAT 157

RESULT 11

US-09-938-842A-3231
; Sequence 3231, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCDP1300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 3231
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-3231

Query Match 3.5%; Score 40.2; DB 9; Length 2000;
Best Local Similarity 45.2%; Pred. No. 2;
Matches 147; Conservative 0; Mismatches 178; Indels 0; Gaps 0;

QY 148 GGACATGTTATGTGAAGGCTTTATGATCAAGAGGTTCCGTTATGATGAAGGGA 207
Db 819 GAAGATGTTTAAGAAAGATCTATGCTGATGAAGATCCAGATTTGTAGTAGCA 878
QY 208 CGTCAAGTCCGGAATTTCACTTCATTTGATTCATCAATTTGGCGGTACGATCT 267
Db 879 GCTTCAGGTTCTGAGATTTCTATTTGGGCTTTAGATTAAACACTCTGTGCTG 938
QY 268 CTGAATCCAGGTGATTTTGTAAACAACAGTGTGTCATTTTCCTTCATTCATATT 327
Db 939 CTGATTCATTTCTCAGTTTCTTAGAGACACGCTTAAAGTTTCGATTTTACTATT 998
QY 328 GTTACCAAGTTGATCGTGCATATGAGTACAGTTCGTTTACATGGAAGCTGATAACA 387
Db 999 ATTCTGACTCCTCAATCTATCTGACTTGAAGATTTTGTGTATTAACAACAATATA 1058
QY 388 GTTAGTGCACAGATTGAGTATGTAATCACAACCTGCTTTCAAACTCAATTTGCCG 447
Db 1059 TGTAGTCAAGATCTACAAATTTCTCAATCATTTCAATAGCTTCTCTCATATATG 1118
QY 448 ATGCCAGTATGCCGTTATGAATTT 472
Db 1119 ATGGTAGAAGAACTGTGTACTT 1143

RESULT 12

US-09-902-941-1883/C
; Sequence 1883, Application US/09902941
; Patent No. US20020172952A1
; GENERAL INFORMATION:
; APPLICANT: Henderson, Robert A.
; APPLICANT: Wang, Tonglong
; APPLICANT: Watanabe, Yoshihiro
; APPLICANT: Johnson, Jeffrey C.

; APPLICANT: Retter, Marc W.
; APPLICANT: Marnerakis, Margarita
; APPLICANT: Carter, Darrick
; APPLICANT: Panger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Bangur, Chaitanya S.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C17
; CURRENT APPLICATION NUMBER: US/09/902,941
; CURRENT FILING DATE: 2001-07-10
; NUMBER OF SEQ ID NOS: 2002
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1883
; LENGTH: 6799
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-902-941-1883

Query Match 3.3%; Score 38; DB 9; Length 6799;
Best Local Similarity 59.1%; Pred. No. 15;
Matches 65; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 235 TTTGATTCATGCAATGTTGGCGGTACACGATCTGTAATCCAGTGGTATTTTGTACA 294
Db 5002 TTTGTTACAGCAAAATCGAGTCCACGAGTCTCAATGCGCATGCTTCTCTGTCA 4943
QY 295 ACAAGTGTGTCATTTGCTTTCATTCATTTATTTGTTACCAAGTTGATCG 344
Db 4942 GTACCTGTGCTACTTATTTCTCTAGATGATTTATTTCTTTATTTATGG 4893

RESULT 13

US-09-849-626-1883/C
; Sequence 1883, Application US/09849626
; Publication No. US20020197669A1
; GENERAL INFORMATION:
; APPLICANT: Bangur, Chaitanya
; APPLICANT: Panger, Gary
; APPLICANT: Wang, Aijun
; APPLICANT: Wang, Tonglong
; APPLICANT: Switzer, Anne
; APPLICANT: McNeill, Patricia
; APPLICANT: Clapper, Jonathan
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C16
; CURRENT APPLICATION NUMBER: US/09/849,626
; CURRENT FILING DATE: 2001-05-03
; NUMBER OF SEQ ID NOS: 1926
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1883
; LENGTH: 6799
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-849-626-1883

Query Match 3.3%; Score 38; DB 9; Length 6799;
Best Local Similarity 59.1%; Pred. No. 15;
Matches 65; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 235 TTTGATTCATGCAATGTTGGCGGTACACGATCTGTAATCCAGTGGTATTTTGTACA 294
Db 5002 TTTGTTACAGCAAAATCGAGTCCACGAGTCTCAATGCGCATGCTTCTCTGTCA 4943
QY 295 ACAAGTGTGTCATTTGCTTTCATTCATTTATTTGTTACCAAGTTGATCG 344
Db 4942 GTACCTGTGCTACTTATTTCTCTAGATGATTTATTTCTTTATTTATGG 4893

RESULT 14

US-10-017-754-1883/C

```
; Sequence 1883, Application US/10017754
; Publication No. US20030054363A1
; GENERAL INFORMATION:
; APPLICANT: Henderson, Robert A.
; APPLICANT: Wang, Tonglong
; APPLICANT: Watanabe, Yoshihiro
; APPLICANT: Johnson, Jeffrey C.
; APPLICANT: Retter, Marc W.
; APPLICANT: Marnerakis, Margarita
; APPLICANT: Carter, Darick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: McNabb, Andria
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C18
; CURRENT APPLICATION NUMBER: US/10/017,754
; CURRENT FILING DATE: 2001-10-29
; NUMBER OF SEQ ID NOS: 2004
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1883
; LENGTH: 6799
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-017-754-1883

Query Match
Best Local Similarity 59.1%; Score 38; DB 9; Length 6799;
Pred. No. 15;
Matches 65; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 235 TTTCATTCAGCATGTTGGCCGACAGCATCTGATCCATCCACGTCGATTTTGTACCA 294
DB 5002 TTTCATTCAGCAAAATGGAGTCCACGAGCTCTCAATGCCATGGTTCTCTGTCA 4943

QY 295 ACAACTGTTCATTCGTTTCATCATCATTTTGTTCACCAAGTTGATCG 344
DB 4942 GTACCTGTGCTGACTTATTTCCTAGATGATTTATTCCTTTATTG 4893

RESULT 15
; Sequence 18737, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
```

```
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 18737
; LENGTH: 510
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL035457.11
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.85
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.9
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 4.6
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.93
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.95
; OTHER INFORMATION: NT HIT: U01287.1, EVALU8 1.30e+00
US-09-864-761-18737

Query Match
Best Local Similarity 46.3%; Score 36.2; DB 10; Length 510;
Pred. No. 10;
Matches 119; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

QY 476 ATGTGACCAACCGGTCACCAAGTTCATTTGCTATCATTTGGTCAGCCAGTTATCAT 535
DB 414 ATGCTGATGACAGCGCTGATCTGATGCAATGATGTTGGTGGACAGATGTGATGAG 355

QY 536 AATGACATGCGATTTCGAACCGTTGATTTCTGCGCGGTTCATTCGCTTGG 595
DB 354 ATGTTGAAGCAGGTGTAAGAAAGTGTGATGACAGGTGACAGATGCTGCTGATG 295

QY 596 TCGATGATGTAACGGTGTACTGTGCAAAATCTAAATGCTGATGATGCTCTTGATA 655
DB 294 ATGCTGATGATGATGATGCTGCTGATGATGATGATGATGATGATGATGATGATG 235

QY 656 AATATTTGCTTAATTAATTTGGAATATATCCACAGATTTAATGCTGCGCCACAGCTCAG 715
DB 234 CAGGTGCTGAAGAAATGTTGTTGATGCAAGTGCATGCTGATGCTGATGCTGATG 175

QY 716 TATCAATATATGCGAT 732
DB 174 CTGATGATGATGCTGCT 158

RESULT 16
; Sequence 5151, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
```

```
;; PRIOR FILING DATE: 2000-10-06
;; PRIOR APPLICATION NUMBER: 60/279,526
;; PRIOR FILING DATE: 2001-03-27
;; NUMBER OF SEQ ID NOS: 8481
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 5151
;; LENGTH: 1103
;; TYPE: DNA
;; ORGANISM: Bacillus clausii
US-09-974-300-5151

Query Match          3.1%; Score 36.2; DB 10; Length 1103;
Best Local Similarity 56.2%; Pred. No. 16;
Matches 68; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 471 TTGATGATGACCAACCGGTCAACCGTTCATTTGCTATCATGTGTCAGCCAGTTTA 530
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 812 TTTCAGGAGATGTCCTTCAGTCTCATTTGCTGATTTTGCTAGCCTTAGTGGCGCAATCGG 871

QY 531 TCATAATGACATGCATTCCTGAACCGTTGATCTTTTCGCCGGTTGTCATTCCTG 590
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 872 TCACGCCGATTTGCAATTTGCTCTGCTTTTGCCGCCCTGTCATCGCACCTG 931

QY 591 C 591
DB 932 C 932

RESULT 17
US-08-781-986A-33
; Sequence 33, Application US/08781986A
; Publication No. US20030054436A1
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5255
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,986A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Benson, Bob
; REGISTRATION NUMBER: 30,446
; REFERENCE/DOCKET NUMBER: PB248PP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 33:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7563 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-781-986A-33

Query Match          3.0%; Score 35.4; DB 7; Length 7563;
Best Local Similarity 53.2%; Pred. No. 79;
Matches 75; Conservative 0; Mismatches 66; Indels 0; Gaps 0;
```

```
QY 16 ATTCCTTTCTACTACACTTAATTCATTCCTTAATTCGATTCGCCGTGCAATGCTGC 75
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5476 ATTGCTTACCGTATGAAGATTAATGAACGTTTTATTATTATACCATGCAATGCGTGA 5535

QY 76 GAAGGTGACCCAGAAATTCATGTGACACACTTCATTAATCAATCAATTTAATACCGT 135
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5536 AAAGTATGCTCGAAGAGCAATATGATTTGGTGGACAAAAGTAACATGTTCAATCGCA 5595

QY 136 AATGCATTCGAGACATGTT 156
DB 5596 CAAGCACGCTTTCGAATGTT 5616

RESULT 18
US-09-901-136-3
; Sequence 3, Application US/09901136
; Publication No. US20030039968A1
; GENERAL INFORMATION:
; APPLICANT: HU, Song et al.
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND
; FILE REFERENCE: C1001273
; CURRENT APPLICATION NUMBER: US/09/901,136
; CURRENT FILING DATE: 2001-07-10
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 378361
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(378361)
; OTHER INFORMATION: n = A,T,C or G
US-09-901-136-3

Query Match          3.0%; Score 35.4; DB 9; Length 378361;
Best Local Similarity 53.2%; Pred. No. 6.2e+02;
Matches 75; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 119 TCATTTTAATACAGTAATGCAATTCGAGACATGTTATGTGAAGGCTTTATGATC 178
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 279626 TCAGTATATAGTTAATATTAATTTTAAAGCATGTGATAGAAAGATTTTCAGT 279685

QY 179 AAGAAGCTGCGGTATGATGAAGGTGACGTCAGTCCGGAATTCACCTTCATTTG 238
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 279686 ATGCAGGTCTCAGTATGCTGGAATAGAAAGAAATACCTGTTTAAACATAAGTAG 279745

QY 239 ATTCAATGAATGTTGCGCTA 259
DB 279746 ATTTCACTGATGCTCAGTGAA 279766

RESULT 19
US-09-864-761-1160/C
; Sequence 1160, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aecm1ca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
```



```

; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 1160
; LENGTH: 465
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC007389.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 5.2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.6
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.89
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.5
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.3
; US-09-864-761-1160

Query Match      3.0%; Score 35.2; DB 10; Length 465;
Best Local Similarity 60.4%; Pred. No. 18;
Matches 58; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 59 CCGTTCACATGCTGTCGAGCGTACGACCAAGATTCGATGCGACCACTTCATACCA 118
DB 295 CCGGGAAAACAGATGAGAGTCTCTCAAAAATTAAGTAACTACGATACGATGACCA 236

QY 119 TCAATTTTATACAGATGATGCATTCGAGACATG 154
DB 235 GCAATTCACACTGCGGTATATATCAAGGAATG 200

RESULT 20
; Sequence 19488, Application us/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
```

```

; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aomico-X-1
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 19488
; LENGTH: 660
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC008125.9
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 6.5
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.9
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 6.5
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 5.8
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 5
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.8
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 3
; OTHER INFORMATION: NT HIT: AF044255.1, EVALUATE 1.10e-01
; US-09-864-761-19488

Query Match      3.0%; Score 35; DB 10; Length 660;
Best Local Similarity 46.2%; Pred. No. 25;
Matches 116; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 461 GTTATGAATTTGGATGTCGACCAACCGGTCAACAGTTTCATTTGCTATCTTGGTC 520
DB 121 GTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 180

QY 521 AGCCAGTTTATCATTAATGACATGCAATTCGTGAAACCGTTGATACCTTTGCGCGGTTG 580
```



```
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 20772
; LENGTH: 583
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL022334.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.3
; OTHER INFORMATION: EST_HUMAN HIT: AW92395.1, EVALUO 9.50e-02
US-09-864-761-20772

Query Match          3.0%; Score 34.8; DB 10; Length 583;
Best Local Similarity 48.1%; Pred. No. 26;
Matches 99; Conservative 0; Mismatches 107; Indels 0; Gaps 0;
```

```
QY 465 TGAATTTTGGATGTTGACCAACCGGTCAACGTTTCATTTGCTATGTCACGCC 524
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 256 TGATGATGGGATGATGATACACTGATGATGATGATGATGATGATGATGATGATG 315
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 525 AGTTATCATTAATGACATGCGATTCGAAACCGTGTGATCTTGGCGGTTGCCA 584
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 316 CGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 375
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 585 TTCTGCTTTGCGATGATGTAACGCTGATCTGTAATTCGTAATGCTGATGATG 644
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 376 TGATGATTAATGATGATGAAGCATGATGATGATGATGATGATGATGATGATG 435
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 645 TGCCTTGCATTAATATTTTCTAAATA 670
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 436 TGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 461
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 23
US-09-815-242-3929
; Sequence 3929, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
```

```
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3929
; LENGTH: 1608
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
US-09-815-242-3929
```

```
Query Match          3.0%; Score 34.8; DB 10; Length 1608;
Best Local Similarity 53.7%; Pred. No. 47;
Matches 72; Conservative 0; Mismatches 62; Indels 0; Gaps 0;
```

```
QY 601 GATGTAACGGTGTACTGTGGAATTCCTAATCTGATGATGCTGCTTGAATAAT 660
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 910 GATCGCATATTTCTGTTGTAAGCACTTAACATGCTGCTTGAATTTGATTCGAT 969
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 661 TTGCTAATTAATTTGATATTCACAGATTTAATGCTGCGCAAGCTCAGCTATAC 720
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 970 ATCGAATTTGATGCTGATTCACAGAAATTAATCTGTAATGTCGTAAGAAATT 1029
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 721 AATATGCGATCG 734
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1030 GGTTCACGATGG 1043
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 24
US-09-815-242-6630
; Sequence 6630, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
```

```

PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
SEQ ID NO 4012
LENGTH: 1959
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AL022334.1
OTHER INFORMATION: EXPRESSED IN BR474, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.1
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.7
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.3
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.7
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.2
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.4
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.2
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.7
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.3
US-09-864-761-4012

Query Match          3.0%; Score 34.8; DB 10; Length 1959;
Best Local Similarity 48.1%; Pred. No. 52;
Matches 99; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

QY 465 TGAATTTGGATGCTGGACCACCGGCACCAAGTTCATTTGCTATCATGTGTCAGCC 524
      ||| | ||||| ||| | | | | | | | | | | | | | | | | | | | | | | |
DB 761 TGATGATGGGAGTATGATGCACACTGATGATGATGGTGATGATGATGATGATGATGA 820
QY 525 AGTTATCATCAAAGAATCGATTCGTAACCGTTGATCTTTCGCCGGTGTCCA 584
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 821 CGGGATCATATGACAGTAGTATGATGATGATGATGATGATGATGATGATGATGACGG 880
QY 585 TTCCTGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 644
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 881 TGATGATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATA 940
QY 645 TGCTCTGATTAATATTTCCTAAATA 670
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 941 TGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATA 966

RESULT 26
US-10-210-296-1
Sequence 1, Application US/10210296
Publication No. US20030021802A1
GENERAL INFORMATION:
APPLICANT: Pfizer Products Inc.
TITLE OF INVENTION: LAWSONIA INTRACELLULARIS PROTEINS, AND RELATED
TITLE OF INVENTION: METHODS
FILE REFERENCE: PC10589A
CURRENT APPLICATION NUMBER: US/10/210,296
CURRENT FILING DATE: 2002-08-01
PRIOR APPLICATION NUMBER: US/09/689,065
PRIOR FILING DATE: 2000-10-12
NUMBER OF SEQ ID NOS: 102
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 6617
TYPE: DNA
ORGANISM: Lawsonia intracellularis
US-10-210-296-1

Query Match          3.0%; Score 34.8; DB 9; Length 6617;
Best Local Similarity 50.6%; Pred. No. 1e+02;
Matches 84; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 87 AGAATTGAATGTGGACCACTTCATTAACAATCAATTTTAATACAGTAATGCAATTGCA 146

```

Db 1490 AGAATTGCTAATGATTTTCAGAACTTAATGATTCATGATCAAGCTGCGTTTGA 1549
QY 147 AGCAGATGTTATGTGAAGCTTTTATGATCAGAAGCTGCCGTATGATGAAGTGG 206
Db 1550 AGAACAAGTTTGTGAAGAAAGCAGTGCAGACGAAGACCAATGCCATGATGAAGATTA 1609
QY 207 ACCTCAAGTTGCCGAATTTTCATTTCCATTTGATTCATGATGATGTT 252
Db 1610 TCTTCGTGCTATTAGAAATATGAAATGCCACGACGACGAGGGAAGT 1655

RESULT 27
US-09-070-927A-217/c
; Sequence 217, Application US/09070927A
; Patent No. US20020120116A1
GENERAL INFORMATION:
APPLICANT: Charles A. Kunsch
Patrick J. Dillon
Steven Barash
TITLE OF INVENTION: Enterococcus faecialis polynucleotides and polypeptides
NUMBER OF SEQUENCES: 982
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/070.927A
FILING DATE: 04-May-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/046,655
FILING DATE: 1997-05-16
APPLICATION NUMBER: 60/044,031
FILING DATE: 1997-05-06
APPLICATION NUMBER: 60/066,009
FILING DATE: 1997-11-14
ATTORNEY/AGENT INFORMATION:
NAME: Kenley K. Hoover
REGISTRATION NUMBER: 40,302
REFERENCE/DOCKET NUMBER: PB369
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 217:
SEQUENCE CHARACTERISTICS:
LENGTH: 8395 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 217:
US-09-070-927A-217

Query Match 3.0%; Score 34.8; DB 10; Length 8395;
Best Local Similarity 53.7%; Pred. No. 1.2e+02;
Matches 72; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 601 GATGCTAACGGTGAATCTGGAATTCCTAATGCTGATGATGCTGCTTGATTAATAT 660
Db 4247 GATCCCTATATTTCTGTGTGAGACATTAAACATGCTGCTGTTGATTTGATTCACAT 4188
QY 661 TTGCTAATAATTTTGGAATACCAACAGATTTAATGCTGCGCAAGAAGCTCAGCTATAC 720
Db 4187 ATCAAAATGTATGGGTGATTCACAAAGAAATTAAGTCTGTAATAATGTGCGAAGAATTT 4128
QY 721 AATATGCGGATCG 734

Db 4127 GGTTCAGCAGATCG 4114

RESULT 28
US-09-070-927A-405
; Sequence 405, Application US/09070927A
; Patent No. US20020120116A1
GENERAL INFORMATION:
APPLICANT: Charles A. Kunsch
Patrick J. Dillon
Steven Barash
TITLE OF INVENTION: Enterococcus faecialis polynucleotides and polypeptides
NUMBER OF SEQUENCES: 982
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/070.927A
FILING DATE: 04-May-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/046,655
FILING DATE: 1997-05-16
APPLICATION NUMBER: 60/044,031
FILING DATE: 1997-05-06
APPLICATION NUMBER: 60/066,009
FILING DATE: 1997-11-14
ATTORNEY/AGENT INFORMATION:
NAME: Kenley K. Hoover
REGISTRATION NUMBER: 40,302
REFERENCE/DOCKET NUMBER: PB369
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 405:
SEQUENCE CHARACTERISTICS:
LENGTH: 1479 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 405:
US-09-070-927A-405

Query Match 3.0%; Score 34.6; DB 10; Length 1479;
Best Local Similarity 44.3%; Pred. No. 50;
Matches 136; Conservative 1; Mismatches 170; Indels 0; Gaps 0;

QY 111 AATAACAATCAATTTTAATACAGTAATGATTCGAGAGACGTGTTATGTGAAGTCT 170
Db 1061 AAGAACAATTTATTTATGATTAATGAGTAATGCAATATGAATGACATGCAAGAAATGT 1120
QY 171 TTATCATCAAGAGTTCGCCGTAATGACGAAGTGCAGTCAAGTTCAGTCCGGAATTTCACT 230
Db 1121 TGAAGCTAAGATGCTGTTATTTATTTATGTAAGTAACTTAGTAAATCGAAGCAGTCTATGC 1180
QY 231 TCCATTTGATTCATGCAATGTTGCGCGTACAGATCTGATCCACGCTGATTTTGT 290
Db 1181 AAAACAAGATCAATATGCCAATGATTTTACCACCGCATGTCGRTTAAGTTAGT 1240
QY 291 AACACAAGCTGTGTGATTTGCTTTCATCATCTTATTTGTAACCAAGTTATGCTGATA 350
Db 1241 ATCAGCTGTGAAAACAATTCAAAAGGCCACAGTGTTTTGAAGCCGTAGTCATCCAGA 1300

RESULT 33
US-10-054-562A-12/c
; Sequence 12, Application US/10054562A
; Patent No. US20020165375A1

```

RESULT 34
US-09-864-761-25339
: Sequence 25339, Application US/09864761
: Patent No. US20020048763A1
: GENERAL INFORMATION:
: APPLICANT: Penn, Sharron G.
: APPLICANT: Rank, David R.
: APPLICANT: Hanzel, David K.
: APPLICANT: Chen, Wensheng
: TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
: TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
: FILE REFERENCE: Aecomica X-1
: CURRENT APPLICATION NUMBER: US/09/864,761
: CURRENT FILING DATE: 2001-05-23
: PRIOR APPLICATION NUMBER: US 60/180,312
: PRIOR FILING DATE: 2000-02-04
: PRIOR APPLICATION NUMBER: US 60/207,456
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: US 09/632,366
: PRIOR FILING DATE: 2000-08-03
: PRIOR APPLICATION NUMBER: GB 24263, 6
: PRIOR FILING DATE: 2000-10-04
: PRIOR APPLICATION NUMBER: US 60/236,359
: PRIOR FILING DATE: 2000-09-27
: PRIOR APPLICATION NUMBER: PCT/US01/00666
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00667
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00664
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00669
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00665
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00668
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00663
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00662
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00661
: PRIOR FILING DATE: 2001-01-30

```

```

; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 25339
; LENGTH: 325
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC023000.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.9
; OTHER INFORMATION: NT HIT: AF164151.1, EVALUE 6.80e-01
; OTHER INFORMATION: EST_HUMAN HIT: BE336733.1, EVALUE 3.20e+00
US-09-864-761-25339
```

```
Query Match          2.9%  Score 34;  DB 10;  Length 325;
Best Local Similarity 56.1%;  Pred. No. 30;
Matches 64;  Conservative 0;  Mismatches 50;  Indels 0;  Gaps 0;
```

```
QY 547 GATTCTGAACCGTTGACTTTCTGCGCGTTCATTCGCTTTGCGATGATGT 606
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 124 GATTATCATATATGTATGATGATGATGATGATGATGATGATGATGATGAT 183
QY 607 AACGGTGAACCTGTGAATTCCTAATGCTGATGATGCTCTTGATAAATAT 660
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 184 GACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 237
```

RESULT 35

```
US-09-864-761-8619
; Sequence 8619, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Hanzel, David R.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeonica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
```

```

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 8619
; LENGTH: 556
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC023000.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.9
US-09-864-761-8619
```

```
Query Match          2.9%  Score 34;  DB 10;  Length 556;
Best Local Similarity 56.1%;  Pred. No. 41;
Matches 64;  Conservative 0;  Mismatches 50;  Indels 0;  Gaps 0;
```

```
QY 547 GATTCTGAACCGTTGACTTTCTGCGCGTTCATTCGCTTTGCGATGATGT 606
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 158 GATTATCATATATGTATGATGATGATGATGATGATGATGATGATGATGAT 217
QY 607 AACGGTGAACCTGTGAATTCCTAATGCTGATGATGCTCTTGATAAATAT 660
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 218 GACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 271
```

RESULT 36

```
US-09-745-288-91
; Sequence 91, Application US/09745288
; Patent No. US20010018058A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Xu, Jianshun
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND
; FILE REFERENCE: 210121.446D1
; CURRENT APPLICATION NUMBER: US/09/745,288
; CURRENT FILING DATE: 2000-12-19
; NUMBER OF SEQ ID NOS: 101
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 91
; LENGTH: 895
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-745-288-91
```

```
Query Match          2.9%  Score 34;  DB 10;  Length 895;
Best Local Similarity 54.9%;  Pred. No. 54;
Matches 67;  Conservative 0;  Mismatches 55;  Indels 0;  Gaps 0;
```

```
QY 282 TATTTTGTAAACACACTGTGTCATTTGCTTCATTCATATATTTGTACCAAGTTGA 341
      |||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 150 TATATCCTTAATCTCAATATATGTCACCTTATTAATAAAGGATTTGCTAACACATGCAT 209
QY 342 TCGTCATATATGATGATCAATGCTTTTACATGGAAGCTGATAAACAGTTAGTGACAGAT 401
```


Db 210 TTGTGACACAGATGCCAAAATTATACATGTAAGTAAATGACCAACCAAGAGATATACAC 269
QY 402 TG 403
Db 270 TG 271

RESULT 37

US-09-764-872-725/C
; Sequence 725, Application US/09764872
; Publication No. US20030050231A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, proteins, and antibodies
; FILE REFERENCE: PA125
; CURRENT APPLICATION NUMBER: US/09/764,872
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PAM or file wrapper
; NUMBER OF SEQ ID NOS: 957
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 725
; LENGTH: 2322
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-872-725

Query Match 2.9%; Score 34; DB 9; Length 2322;
Best Local Similarity 54.9%; Pred. No. 94;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 282 TATTTTGTACACACACGTGTGTCATTGCTTCATCCATTTATTTTACCAAGTTGA 341
Db 2179 TATATCTTACTTCAATTTATTTGACCTTATTAATTAAGGTGATTTCTTAACACATGCAT 2120
QY 342 TCGTGATATCGAGTACAAATGCTTTTACATGGAAGCTGATAAACAGTTAGTCACAGAT 401
Db 2119 TTGTGACACAGATGCCAAAATTATACATGTAAGTAAATGACCAACCAAGAGATATACAC 2060

QY 402 TG 403
Db 2059 TG 2058

RESULT 38

US-08-781-986A-272/C
; Sequence 272, Application US/08781986A
; Publication No. US20030054436A1
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5255
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,986A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Benson, Bob

; REGISTRATION NUMBER: 30,446
; REFERENCE/DOCKET NUMBER: PB248PP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 272:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6035 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-781-986A-272

Query Match 2.9%; Score 34; DB 7; Length 6035;
Best Local Similarity 50.6%; Pred. No. 1.6e+02;
Matches 82; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 237 TGATTATGCAATGTTGCGGTACACGATCTCTGAATCCAGTGTATTTTGTGAACAAC 296
Db 5958 TTGTATGCAATGATCGCGCTTAACGTCACAATGATATTTTATAGTTGTACAGC 5899
QY 297 AACTGTGTCAATTCGTTTATCCATTTATTTGTACCAAGTTGATGCTATTCGAGT 356
Db 5898 AGCATTTAGATATTAATTTATTTACGTTCACTGTTACAAAGAAATTTATTTGCTGTCCC 5839
QY 357 ACATGCTTTTACATGGAAGCTGATTAACAGTTAGTGCACA 398
Db 5838 TGCGCGCTTTGACATATACGACCTCAAAATTAATTCAAGA 5797

RESULT 39

US-09-815-242-4340
; Sequence 4340, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: EUTRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4340
; LENGTH: 1170
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-815-242-4340

Query Match 2.9%; Score 33.8; DB 10; Length 1170;
Best Local Similarity 52.5%; Pred. No. 71;
Matches 74; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

Search completed: March 28, 2003, 17:58:31
Job time : 529 secs

```
QY 16 ATTGCTTTCTGTACTACACTTATTCATGTCTTATTCGATCCGGTTGACAAATGGTGTC 75
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 796 ATTGCTTACCGTATGAAGATATGAAAGCTTTTATTATTACCGATCAATGCGTGCA 855
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 76 GAAGGTGAGCCGAATTAATGATGGACCACTTCATACAAATCAATTTAATACAGT 135
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 856 AAAGGTATGCCGTGAAGAGATATGATTGGGTGTCAAAAGTAAGTCAATGTCATCGCAA 915
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 136 AATGCATTCGAGACATGTT 156
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 916 CAAGCACGCTTTCGCAAAATGTT 936
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 40

```
US-09-815-242-8204
: Sequence 8204, Application US/09815242
: Patent No. US20020061569A1
: GENERAL INFORMATION:
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsen, Karl L.
: APPLICANT: Zyskind, Judith W.
: APPLICANT: Wall, Daniel
: APPLICANT: Twick, John D.
: APPLICANT: Cair, Grant J.
: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: Identification of Essential Genes in
: TITLE OF INVENTION: Prokaryotes
: FILE REFERENCE: ELITRA.011A
: CURRENT APPLICATION NUMBER: US/09/815,242
: CURRENT FILING DATE: 2001-03-21
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: NUMBER OF SEQ ID NOS: 14110
: SOFTWARE: PastedSeq for Windows Version 4.0
: SEQ ID NO 8204
: LENGTH: 1182
: TYPE: DNA
: ORGANISM: Staphylococcus aureus
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (1)...(1182)
US-09-815-242-8204
```

Query Match 2.9%; Score 33.8; DB 10; Length 1182;
Best Local Similarity 52.5%; Pred. NO. 72;
Matches 74; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

```
QY 16 ATTGCTTTCTGTACTACACTTATTCATGTCTTATTCGATCCGGTTGACAAATGGTGTC 75
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 805 ATTGCTTACCGTATGAAGATATGAAAGCTTTTATTATTACCGATCAATGCGTGCA 864
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 76 GAAGGTGAGCCGAATTAATGATGGACCACTTCATACAAATCAATTTAATACAGT 135
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 865 AAAGGTATGCCGTGAAGAGATATGATTGGGTGTCAAAAGTAAGTCAATGTCATCGCAA 924
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 136 AATGCATTCGAGACATGTT 156
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 925 CAAGCACGCTTTCGCAAAATGTT 945
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2003, 15:14:55 ; Search time 1642 Seconds
(without alignments)
11451.255 Million cell updates/sec

Title: US-10-054-562a-3

Perfect score: 1161
Sequence: 1 atgatgatctctatctatgc.....ttcgtccaaatcagaagca 1161

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_iny:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	481.8	41.5	721	12	BG310598	SMOV3MCA1
2	445.2	38.3	638	12	BF482157	SMOV3MCA1
3	432.6	37.3	775	12	BF824707	SMOV3MCA1
4	403	34.7	758	12	BF918213	SMOV3MCA1
5	392.6	33.8	647	12	BF727586	SMOV3MCA1
6	390.2	33.6	665	9	A1111196	SMOV3MCA1

7	387.2	33.4	740	12	BF727527	SMOV3MCA1
8	385	33.2	648	12	BF400424	SMOV3MCA1
9	376.8	32.5	646	13	BM285148	SMOV3MCA1
10	376.2	32.4	747	12	BG809012	SMOV3MCA1
11	372.2	32.1	628	9	AA701731	SMOV3MCA1
12	370.2	31.9	644	12	BF400417	SMOV3MCA1
13	369.8	31.9	774	12	BF599177	SMOV3MCA1
14	364.2	31.4	672	9	AA618895	SMOV3MCA1
15	363.6	31.3	612	13	BM284858	SMOV3MCA1
16	358.6	30.9	639	9	A1322117	SMOV3MCA1
17	355.8	30.6	610	9	AA668071	SMOV3MCA1
18	351.6	30.3	648	13	BM284957	SMOV3MCA1
19	336.8	29.0	635	12	BF482126	SMOV3MCA1
20	333.8	28.8	624	13	BM284979	SMOV3MCA1
21	331.4	28.5	554	12	BF482099	SMOV3MCA1
22	324.2	27.9	776	9	AA585626	SMOV3MCA1
23	312.4	26.9	537	9	AA668051	SMOV3MCA1
24	307	26.4	691	12	BF727629	SMOV3MCA1
25	306.8	25.8	552	9	A1322068	SMOV3MCA1
26	299.6	25.8	615	9	A1317885	SMOV3MCA1
27	299	25.6	543	12	BF758436	SMOV3MCA1
28	297.6	25.6	694	10	AA355791	SMOV3MCA1
29	292.6	25.2	534	13	BI501509	SMOV3MCA1
30	291.2	25.1	592	13	BI783374	SMOV3MCA1
31	285	24.5	591	13	BF400320	SMOV3MCA1
32	280.4	24.2	622	12	BF400320	SMOV3MCA1
33	278.8	24.0	363	9	AA625020	SMOV3MCA1
34	277.8	23.9	558	10	AA313040	SMOV3MCA1
35	275.6	23.7	543	10	AA312995	SMOV3MCA1
36	274.6	23.7	577	10	AA288114	SMOV3MCA1
37	269.2	23.2	354	9	AA841200	SMOV3MCA1
38	269.2	23.2	705	12	BF918256	SMOV3MCA1
39	267.8	23.1	466	9	AA625024	SMOV3MCA1
40	263	22.7	459	9	AA618952	SMOV3MCA1
41	260.4	22.4	395	9	AA625010	SMOV3MCA1
42	257.6	22.2	523	13	BI781683	SMOV3MCA1
43	257	22.1	369	9	AA625022	SMOV3MCA1
44	253.6	21.8	492	10	AA313041	SMOV3MCA1
45	249	21.4	541	13	BI662949	SMOV3MCA1

ALIGNMENTS

RESULT 1
BG310598
LOCUS 721 bp mRNA linear EST 23-FEB-2001
DEFINITION SMOV3MCA15F02SK Onchocerca volvulus molting L3 larva cDNA
(SL96MLM-Ovml3) Onchocerca volvulus cDNA clone SMOV3MCA15F02 5',
mRNA sequence.

ACCESSION BG310598
VERSION BG310598.1 GI:13112315
KEYWORDS
SOURCE Onchocerca volvulus.
ORGANISM Onchocerca volvulus

REFERENCE 1 (bases 1 to 721)
AUTHORS Williams, S.A., Lizotte-Waniewski, M., Ianey, S. and Lustigman, S.
TITLE Genes expressed in molting L3 larvae of Onchocerca volvulus
JOURNAL Unpublished (1997)
COMMENT Contact: Steven A. Williams
Molecular Parasitology
Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith
College, Northampton, MA, 01063, USA
Tel: 4135853826
Fax: 4135853786
Email: genome@smith.edu

FEATURES
source
1..721
/organism="Onchocerca volvulus"

```

BASE COUNT
ORIGIN
213 a      140 C      158 g      210 t
      /strain="Kumba, Cameroon"
      /db_xref="taxon:6282"
      /clone="SMOV3MCAM55F02"
      /clone.lib="Onchocerca volvulus molting L3 larva cDNA
      (SL96MLM-Ovml3)"
      /dev_stage="molting L3"
      /lab_host="XLI-Blue MRF"
      /note="Vector: lambda uni-zap XR; Site.1: Eco RI; Site.2:
      Xho I; Filarial nematode parasite of humans. Third-stage
      larvae, L3, were isolated from infected black flies in
      Cameroon (forest strain). The L3 were cultured in 20% FCS
      in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in
      culture. L3 of O. volvulus molt to fourth-stage larvae by
      day 5 in culture. mRNA was isolated from approximately
      6000 molting larvae (ml3), 2000 larvae from day 1, 2 or 3
      in culture, and converted to double-stranded cDNA using
      reverse transcriptase and oligo(dT) followed by RNase H
      and DNA pol I. The library was constructed in the lambda
      Uni-Zap XR vector and has 1 x 10E6 independent
      recombinants and the average insert size is ~1200 bp. The
      library was constructed by Sara Lustigman and Michelle
      Lizotte-Maniewski in the laboratory of Dr. S. A. Williams
      The library is available from Dr. Sara Lustigman (email:
      slustigman@mc.org)."

```

Query Match	41.5%	Score 481.8;	DB 12;	Length 721;
Best Local Similarity	85.4%;	Pred. No. 1.5e-123;		
Best 537; Conservative	0;	Mismatches 92;	Indels 0;	Gaps 0;

QY	160	GTAAGAGCTCTTTATGATCAAGACAGCTGGCCCTAATGATGAAGGAGGAGCGTCAAGCTGGCC	219
Db	1	GTAAGAGCTTATACGATCAGCGACAGAGATGCCAAATGATGAAGGTGGACGTAGGTAGGCC	60
QY	220	GGAATTTCACTTCCATTTGATTCATGACAAATGTTGGCGGTACACGATCTCTGATCCACGT	279
Db	61	GGAATTTCACTTCCCTTTGATTCCTCGTAAATGAGCACGACACGTTCTGTAATACACGT	120
QY	280	GGAATTTTGTATACCAACAAGCTGTGCATTTGCGTTCAATTCATTTTGTATCCCAAGTT	339
Db	121	GGAATTTTGTATACCACTCAGTTGTTTTCATCTTCATTCATTCACCTTTGTTGTGCAAAAGTT	180
QY	340	GATCGTGCAATACGAGTACCAATGCTTTTACATGGAAGCTGATAAAAAGATTAGTCACAG	399
Db	181	GATCGAGCATATCGAATACCAATGCTTTTACATGGAAGCTGATAAAGCATGATAGCGCTCAA	240
QY	400	ATTGGGATCTGGAATACAAACAGCTCTTTCAACATCAATTTGCCATGCGCATATGC	459
Db	241	CTTGAAGTTTCCGAATATACAACTGCTATTTGCACACAAATTTGATCCGATGCTGTATGC	300
QY	460	CGTATGAAATTTTGGATGGTGGACCAACCGGTCAACAGTTCAATTTGCTATCATATGGT	519
Db	301	CGATATGAGATTTTGATGGTGGACATCCGGACAACTGTTCAATTTTGCTAACATTTGTT	360
QY	520	CAGCAGTTTATCAATAATGAGCATGCGATTTCTGAACCCGTTGATCTTTCTGCGCGTT	579
Db	361	CAACCGAGTACCAATAATGAGCATGATATTTCTGAACAGTTGATATCTCTGCTCTTT	420
QY	580	GTTCACTTCTCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	639
Db	421	GTAATTCATCTCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	480
QY	640	GGATGTGCTCTTGATTAATAATTTGCTATAAATTTGGATAATCCAAAGATTTAATGGCT	699
Db	481	GGTGTGCTTTGGACAATAATTTGCTACAACTTGGATAATCCAAAGATTTAATGGCT	540
QY	700	GGCCAAAGAGCTCAGATTAACAATAATGCGATGATGATGATGATGATGATGATGATGATGATGAT	759
Db	541	GGACAAAGAGCTCAGCTTACAAATAATGCGATGATGATGATGATGATGATGATGATGATGATGAT	600
QY	760	ATCAGTATTACATTAAGAACAATAAG	788

DB 601 ATTGACTATTACAAATTAAAGCAACCAACACAG 629

RESULT 2
BF482157
LOCUS
DEFINITION
638 bp mRNA linear EST 06-DEC-2000
SMOV3MCAM48D12SK Onchocerca volvulus molting L3 larva cDNA
(SL1956MLM-Ovm13) Onchocerca volvulus cDNA clone SMOV3MCAM48D12 5',
mRNA sequence.
BF482157
BF482157.1 GI:11565458
EST.
Onchocerca volvulus.
Onchocerca volvulus
Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
Onchocercidae; Onchocerca.
1 (bases 1 to 638)
Williams, S.A., Lizotte-Waniewski, M., Laney, S. and Lustigman, S.
Genes expressed in molting L3 larvae of Onchocerca volvulus
unpublished (1997)
Contact: Steven A. Williams
Molecular Parasitology
Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith
College, Northampton, MA, 01063, USA
Tel: 4135853826
Fax: 4135853786
Email: genome@smith.edu
Seq method: pluscript SK.
Location/Qualifiers

```

/organism="Onchocerca volvulus"
/strain="Kumba, Cameroons"
/db_xref="taxon:6282"
/cclone="SMOV3MCAM48D12"
/clone_1id="Onchocerca volvulus molting L3 larva cDNA
(SL96MIM-Ovml3)"
/dev_stage="molting L3"
/lab_host="XLI-Blue MRF"
/note="Vector: Lambda uni-ZAP XR; Site_1: Eco RI; Site_2:
Xho I; Filarial nematode parasite of humans. Third-stage
larvae, L3, were isolated from infected black flies in
Cameroon (forest strain). The L3 were cultured in 20% FCS
in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in
culture. L3 of O. volvulus molt to fourth-stage larvae by
day 5 in culture. mRNA was isolated from approximately
6000 molting larvae (mL3), 2000 larvae from day 1, 2 or 3
in culture, and converted to double-stranded cDNA using
reverse transcriptase and oligo(dt) followed by RNase H
and DNA pol I. The library was constructed in the lambda
uni-Zap XR vector and has 1 x 10E6 independent
recombinants and the average insert size is ~1200 bp. The
library was constructed by Sara Lustigman and Michelle
Licotte-Maniewski in the laboratory of Dr. S. A. Williams.
The library is available from Dr. Sara Lustigman (email:
slustl@nybc.org)."

```

	BASE COUNT	188 a	123 c	140 g	187 t
ORIGIN					
Query Match	38.3%; Score 445.2; DB 12;				Length 638;
Best Local Similarity	83.1%; Pred. No. 2.5e-113;				
Matches 507; Conservative	0; Mismatches 103; Indels				0; Gaps 0;
QY	257 GTACAGCATCTCTGAATCCAGCGTGTATTTTGTAAACAACAACGTTGTCAATTGGTTTC				316
Dd	3 GCACGAGTGTTGGTAATAATCCCGTGTATTTTGTGCATCATCGTAGTTGTGCATCATTC				62
OY	317 ATTCATTATTTGTACCACAAATGTTATCGTGCATATCCAGTACATGCTTTACATGGAA				376
Dd	63 ATTCATTGTTTGTACAAAAGTTGATGCAGCAATTCGAATTCATGCTTTTACATGGAA				122
OY	377 CTGTAATAAAGAGTAGTCAGAGATTCAGATATCTGGAATCACAAACGCTTTTCAAATC				436

RESULT 3
BF824707 LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BE824707 775 bp mRNA linear EST 13-JAN-2001
SW63MCAM52G1SK Onchocerca volvulus molting L3 larva cDNA
(S196MIM-Oym3) Onchocerca volvulus CDNA clone SW63MCAM52G11 5',
mRNA sequence.
BE824707
BE824707.1 GI:12166567
EST.
Onchocerca volvulus.
Onchocerca volvulus
Euarthropoda; Metazoa; Nematoda; Chromadorea; Splturida; Filarioidea;
Onchoceridae; Onchocerca.
1 (bases 1 to 775)
Williams, S.A., Lizotte-Waniewski, M., Laney, S. and Lustigman, S.
Genes expressed in molting L3 larvae of Onchocerca volvulus
Unpublished (1997)
Contact: Steven A. Williams
Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith
College, Northampton, MA, 01063, USA
Tel.: 4135853826
Fax: 4135853786
Email: genome@smith.edu
Seq primer: pBluescript SK.
Location/Qualifiers
1..775

```

/organism="Onchocerca volvulus"
/strain="Kumba, Cameroons"
/db_xref="taxon:6282"
/clone="SMOV3MCAM52G11"
/clone_1lb="Onchocerca volvulus molting L3 larva cDNA
(SL96MLM-Ovml3)"
/dev_stage="molting L3"
/lab_host="XLI-Blue MRF"
/notes="Vector: Lambda uni-ZAP XR; Site_1: Eco RI; Site_2:
Not I; Filarial nematode parasite of humans. Third-stage

```

BASE COUNT	215	a	144	C	166	G	250	T
Query Match	37.3%							
Best Local Similarity	82.4%							
Matches	556							
Conservative	0							
Mismatches	114							
Indels	5							
Gaps	5							
QY	50	ATTGATTC	CCGGTTG	GACAA	TGCTGT	CGAAG	GTGAC	CCAGAA
DB	106	ATGCATAT	CCGGTTG	ATTAAC	GGTGTAG	AAAGG	TGAAC	CGAAAT
QY	110	CAATACAA	TCGAATTT	TANTAC	GTANT	GCATTT	GGAGN	ACATGTT
DB	166	CAATACAA	TCGAATTT	TANTAC	GTANT	GCATTT	GGAGN	ACATGTT
QY	170	TTTATGAT	CAAGAA	GGTTGCC	GTATGAT	GAAAG	GTGAC	AGTCGCA
DB	226	TATAGATC	ACGAGC	AGATGCC	GAATGAT	GAAGT	GAAGT	AGTGAAC
QY	230	TTCCATTT	GATTCAT	GCACAT	GTGTG	CGCGT	TACAC	GCATCTC
DB	286	TTCCATTT	GATTCAT	GCACAT	GTGTG	CGCGT	TACAC	GCATCTC
QY	290	TAAACAA	CAAGTGT	CTCATTT	CGTTG	TCATTC	CAATTT	TGTTGTT
DB	346	TCACATC	AGTTGTT	CTCATTT	CGTTG	TCATTC	CAATTT	TGTTGTT
QY	350	ATCGATCA	ATTCGTT	TTCATTC	ATGAG	CGCTG	ATAAAC	AGTTAGT
DB	406	ATCGATCA	ATTCGTT	TTCATTC	ATGAG	CGCTG	ATAAAC	AGTTAGT
QY	410	CTGAATC	ACAACTG	CTTTT	CAACT	CAAAAT	TGTC	CCGATG
DB	466	CCGAATC	ACAACTG	CTTTT	CAACT	CAAAAT	TGTC	CCGATG
QY	470	TTTTGAT	GGTGG	ACCAAC	CGGTG	ACACAG	TTGCT	ATTTGCT
DB	525	TTCTTGA	TGGTGG	ACCAAC	CGGTG	ACACAG	TTGCT	ATTTGCT
QY	530	ATCATTA	TGACAT	CGCATTT	CTGAAC	CGGTG	ATTAAT	TGTCAT
DB	585	-CCATAA	TGACAT	CGCATTT	CTGAAC	CGGTG	ATTAAT	TGTCAT
QY	590	GCTTTG	ATGATG	ATGAC	GGTGT	ATCTG	TGGAAT	TCTAA
DB	643	GCTTTG	ATGATG	ATGAC	GGTGT	ATCTG	TGGAAT	TCTAA
QY	650	TTTGATTA	ATTTG	CTAATA	ATTTG	GAATA	ATTCGA	AGATTA
DB	702	TGGGCA	CAATTTT	GCTCA	ACAC	CTGG	AAATTC	-ACTG
QY	710	CTCAAG	TATCA	AAAT	724			
DB	761	CTCAAG	TATCA	AAAT	775			

```

LOCUS       BP918213              758 bp    mRNA    linear    EST 19-JAN-2001
DEFINITION   SMOV3MCA53F10SK Onchocerca volvulus molting L3 larva cDNA
              (SL96MLM-OvML3) Onchocerca volvulus cDNA clone SMOV3MCA53F10 5',
              mRNA sequence.
ACCESSION    BF918213
VERSION      BF918213.1  GI:12313988
KEYWORDS     EST.
SOURCE       Onchocerca volvulus.
              Onchocerca volvulus
ORGANISM     Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
              Onchocercidae; Onchocerca.
REFERENCE    1 (bases 1 to 758)
AUTHORS      Williams,S.A., Lizotte-Waniewski,M., Laney,S. and Lustigman,S.
TITLE        Genes expressed in molting L3 larvae of Onchocerca volvulus
JOURNAL      Unpublished (1997)
COMMENT      Contact: Steven A. Williams
              Molecular Parasitology
              Smith College Department of Biological Sciences
              Department of Biological Sciences, Clark Science Center, Smith
              College, Northampton, MA, 01063, USA
              Tel: 4135853826
              Fax: 4135853786
              Email: genom@smith.edu
              Seq primer: plusescript SK.
FEATURES     source
              1..758
               location/Qualifiers
               /organism="Onchocerca volvulus"
               /strain="Kumba, Cameroons"
               /db_xref="taxon:6282"
               /clone="SMOV3MCA53F10"
               /clone_1lb="Onchocerca volvulus molting L3 larva cDNA
               (SL96MLM-OvML3)"
               /dev_stage="molting L3"
               /lab_host="XLI-Blue MRF"
               /note="Vector: Lambda uni-ZAP XR; Site.1: Eco RI; Site.2:
               Xho I; Filarial nematode parasite of humans. Third-stage
               larvae, L3, were isolated from infected black flies in
               Cameroon (forest strain). The L3 were cultured in 20% FCS
               in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in
               culture. L3 of O. volvulus molt to fourth-stage larvae by
               day 5 in culture. mRNA was isolated from approximately
               6000 molting larvae (mL3), 2000 larvae from day 1, 2 or 3
               in culture, and converted to double-stranded cDNA using
               reverse transcriptase and oligo(dT) followed by Rnase H
               and DNA pol I. The library was constructed in the lambda
               uni-Zap XR vector and has 1 x 10E6 independent
               recombinants and the average insert size is ~1200 bp. The
               library was constructed by Sara Lustigman and Michelle
               Lizotte-Waniewski in the Laboratory of Dr. S. A. Williams.
               The library is available from Dr. Sara Lustigman (email:
               slustig@nhybc.org)."
BASE COUNT   212 a      139 c      160 g      247 t
ORIGIN
Query Match      34.7%; Score 403; DB 12; Length 758;
Best Local Similarity 81.4%; Pred. No. 1.6e-101;
Matches 528; Conservative 0; Mismatches 115; Indels 6; Gaps 5;
QY      50 ATTGATTCGGGTGATGCGATGCGAGTCAGACGAGAAATGGAATGGACCACTT 109
      11 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      105 ATGCTATTCCGGTGTATACGCGTGAAGGTGACACGAAATGGAATGGTCCACACT 164
      11 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      110 CAATAACAATCAATTTTAATACACGTAATGATTCGAGAGACATGTTATGTGAAGGTC 169
      11 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      165 CAATTACTGTCACACTTAACTCGTAATCCATTGAAGACATGTAATCGTGAAGGCT 224
      11 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      170 TTATGATCAAGAAGTCCGTAATGATGAGAGCGAGCGTAAGTTCGCGGAATTTCC 229
      11 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      225 TATACGATAGGACAGATGCGCAATGATGAGGTGAGAGTACGAGCCGGAATTTGAC 284
      11 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      230 TTCCATTTGATTCATGCAATGTTGGCGGTCACGATCTCTGATCCACCGTGTATTTTGG 289
      11 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      285 TTCCGTTGATTCCTGTATATGACACGTAACGTTCTTAATTCACCGTGTATTTTGG 344

```

```

QY      290 TAACAACACATGTTGTCATATTTTCGTTTCAATCCATTAATTTGTACCAAAAGTTGATCGTCA 349
      11 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      345 TCACATCAGTGTGTGCATCTCATTCACATCCATTAATTTGTGACAAAAGTTGATCGAGCAT 404
      11 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      350 ATGAGTACATACGCTTTTACATGAGAGCTGATTAACAGTAGTAGTCACAGATTAGAGCTAT 409
      11 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      405 ATGAAATACAAAGCTTTTACATGAGAGCTGATTAAGACATTTAGCGCTGATGAAGTTT 464
      11 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      410 CTGAATTCACAACTGCTTTTCAACATCAAAATTTGCCAGTACGATG-CCGTTATGAA 468
      11 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      465 CCGAAATGACAACTGATTTTGCACACAAATTTTACCGATGCTGTATGCCGATATGAG 524
      11 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      469 ATTTTGGATGGTGACCAACCGGTCAACAGATTCATTAATTTGCTATATGTCAGCCAGTT 528
      11 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      525 ATTCGTGATGAGGACCAATCCGACAAACCGTTCATTTGTCACATGTCATCAACAGT 584
      11 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      529 TATCATTAATGACATGCAATCTCTGAACCGTGTGATTAATTTTTCGCGGTTTCCATTCC 588
      11 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      585 TACCATTAATGACATGATGATCTGTAACAG-TGATACATTTCTGTGCTTGTAC--ATTG 641
      11 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      589 TGCTTTGATGATGATGCTAACGCTGATCTGAGGAAATCTAAATGCTGATGATGTGCT 648
      11 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      642 ATGCTGTTGATGATGATGATGATAC-GTGATTAAGGTGGAATTTTAATGCTGATGG-TGTGCT 699
      11 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      649 CTTGATTAATTAATTTTCTTAATTAATTTGGAATTCACACAGATTTAATG 697
      11 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      700 TTGGACAATTTTTCCTCACAACTTGGGATATCCACACATGATTTATGG 748
      11 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
RESULT 5
BF727586
LOCUS       BF727586              647 bp    mRNA    linear    EST 08-JAN-2001
DEFINITION   SMOV3MCA51C03SK Onchocerca volvulus molting L3 larva cDNA
              (SL96MLM-OvML3) Onchocerca volvulus cDNA clone SMOV3MCA51C03 5',
              mRNA sequence.
ACCESSION    BF727586
VERSION      BF727586.1  GI:12045447
KEYWORDS     EST.
SOURCE       Onchocerca volvulus.
              Onchocerca volvulus
ORGANISM     Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
              Onchocercidae; Onchocerca.
REFERENCE    1 (bases 1 to 647)
AUTHORS      Williams,S.A., Lizotte-Waniewski,M., Laney,S. and Lustigman,S.
TITLE        Genes expressed in molting L3 larvae of Onchocerca volvulus
JOURNAL      Unpublished (1997)
COMMENT      Contact: Steven A. Williams
              Molecular Parasitology
              Smith College Department of Biological Sciences
              Department of Biological Sciences, Clark Science Center, Smith
              College, Northampton, MA, 01063, USA
              Tel: 4135853826
              Fax: 4135853786
              Email: genom@smith.edu
              Seq primer: plusescript SK.
FEATURES     source
              1..647
               location/Qualifiers
               /organism="Onchocerca volvulus"
               /strain="Kumba, Cameroons"
               /db_xref="taxon:6282"
               /clone="SMOV3MCA51C03"
               /clone_1lb="Onchocerca volvulus molting L3 larva cDNA
               (SL96MLM-OvML3)"
               /dev_stage="molting L3"
               /lab_host="XLI-Blue MRF"
               /note="Vector: Lambda uni-ZAP XR; Site.1: Eco RI; Site.2:
               Xho I; Filarial nematode parasite of humans. Third-stage
               larvae, L3, were isolated from infected black flies in
               Cameroon (forest strain). The L3 were cultured in 20% FCS
               in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in
               culture. L3 of O. volvulus molt to fourth-stage larvae by
               day 5 in culture. mRNA was isolated from approximately

```

6000 molting larvae (tm3), 2000 larvae from day 1, 2 or 3 in culture, and converted to double-stranded cDNA using reverse transcriptase and oligo(dT) followed by RNase H and DNA pol I. The library was constructed in the lambda Uni-Zap XR vector and has 1 x 10⁶ independent recombinants and the average insert size is ~1200 bp. The library was constructed by Sara Lustigman and Michelle Lizotte-Waniewski in the laboratory of Dr. S. A. Williams. The library is available from Dr. Sara Lustigman (email: slustigman@bc.org).

BASE COUNT	186 a	144 c	134 g
ORIGIN			

Query Match	33.8%;	Score 392.6;	DB 12;	Length 647;
Best Local Similarity	84.8%;	Pred. No. 1.2e-98;		
Matches 452;	Conservative 0;	Mismatches 79;	Indels 2;	Gaps 1;

9y 1 ATGAGATTGCGCTTATTGCTTCTGTACTACACTATTGCATTTGCTATTTGCCG 60
 ||||| ||||| ||||| ||| ||| ||||| ||||| |||||
 Db 84 ATGATGCTTCGCTTATTCGCTTCTGTGACACTGGCGCTTTTGCATTTGTCTAATTCCT 144

Dy 61 GTTGCATATGGTGCTGGAAGGTGACCGAATAATTGAATGTGGACCAACTTCATATAACATC 12
||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
Db 144 GTTGACAACGGGTCTCGAAGGAGAGCCAGAATTGAGTGGCGGACCACCTTCGATAAATATC 20

Db 204 AATTTTAAACAGCATGCATTGCGAGGACATGTCTATGTGAAGGCTTTACATCAA 26

C7 101 GAGGCTATCCGTAAAGTGAAGGTCACAGTGGCGGAAATTCATTTCATTTGAT 24
| | | | |
Db 264 GCAGCTTGCCGTAATGATGAAGGTGGACGTCAGTGGCCGGAATTCCTCCCTCGAT 32

324 TCTTGCATGTTGCACCGCACGATCTTGAATCCGGAGTATCTTGTAAACAACCTACT 38

Db 384 GTTGATATCTCCTTCACCCACITGTTGTGACAAAAGTTGATGTGATACCGAGTACAG 444

Db 444 TGCTTTACATGGAAGCTGACAAACGGTGCACCCAGATCGAAGTGTCCGAATCACA 50

Db 504 ACTGCTTTCAAACTCAGATTCTTCATGCTGTTGGCCGATATGAATCTTGATGGT 56

Db 564 GGACACTGGACAAACCAATTCACTGGCTACCATTCGGGCGACGACAGTTTAT 616

LOCUS	665 bp	mrna	linear	EST 31-AU
DEFINITION	SMOVMCA1232sk Onchocerca volvulus molting 13 larva cdna			
RESULT 6				
AI111196				
AI111196				

(S596MLM-CVMLJ) Onchocerca volvulus cDNA clone SMOV3MCA1232 5'
 mRNA sequence.
 A1111196
 A1111196.1 GI:3510080
 ACCESSION
 VERSION

KEYWORDS	EST.
SOURCE	<i>Onchocerca volvulus</i> .
ORGANISM	<i>Onchocerca volvulus</i>
Eukaryota:	Metazoa:
	Nematoda:
	Chromadorea:
	Solirurida:
	Eliairid

REFERENCE AUTHORS TITLE	1 (bases 1 to 665) Williams, S.A., Lizotte-Waniewski, M., Laney, S. and Lustigman, S. Genes expressed in molting I/3 larvae of <i>Onchocerca volvulus</i>
Onchocercidae; Onchocerca.	

JOURNAL
UNPUBLISHED (1991)
COMMENT
Contact: Steven A. Williams
Molecular Parasitology

Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith
College, Northampton, MA, 01063, USA
Tel: 413/5853826
Fax: 413/5853786

Email: genome@smith.edu
 Seq primer: pbluescript SK.
 Location/Qualifiers
 1. .665
 /organism="Orchocerca volutus"

```

/organism="Onchocerca volvulus"
/strain="Kumba, Cameroons"
/db_xref="taxon:6282"
/clone="SMOV3MCA1232"
/clone_lib="Onchocerca volvulus molting L3 larva cDNA

```

```

/dev_stage="molting L3"
/lab_host="XLI-Blue MRP"
/notes=Vector: lambda uni-zab XR: site_1: Eco RI: Sitt
Xho I; Filarial nematode parasite of humans. Third-st

```

Cameroon (forest strain). The L3 were cultured in 208 in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in culture. L3 of *O. volvulus* molt to fourth-stage larvae day 5 in culture. mRNA was isolated from approximately

in culture, and converted to double-stranded cDNA using reverse transcriptase and oligo(dT) followed by RNasease and DNA pol I. The library was constructed in the λ Uni-Zap XR vector and has 1 x 10⁶ independent

Library was constructed by Sara Lustigman and Michell Lisotte-Waniewski in the laboratory of Dr. S. A. Will. The library is available from Dr. Sara Lustigman (email: slustigmen@ybc.org).

BASE COUNT	186 a	129 c	136 g	213 t	1 others
ORIGIN					
Query Match	33.6%	Score 390.2;	DB 9;	Length 665;	

Best Local Similarity 84.1%; Pred. No. 5,8e-98;
Matches 451; Conservative 0; Mismatches 84; Indels 1; Gaps
0
50 ATTGGATTCCGGTACACATGGTGTGCAAGGTGACCGAGAAATTGAATGGACAACTT 109

[illegible]

Oy 170 TTTATGATCAAGAAAGTTGCCGTATATGATGAAGTGCACGTCGAAGTTGCCGGAATTTCAC 229
+ | | | | | | | | | | | | | | | | | | | | |
Db 237 TATGCGATCAGCAGGAGATCCGGAATATGATGAAGTGGACGTCAGCTAGCGGGAATTTCAC 266

Dc	TTTCATTTGATTCAATGCAGTCTGGCGGTACAGCATCTCGAATCCACAGGATTTTTC	289
Dy	TTTCATTTGATTCAATGCAGTCTGGCGGTACAGCATCTCGAATCCACAGGATTTTTC	290
Dd	TTTCATTTGATTCAATGCAGTCTGGCGGTACAGCATCTCGAATCCACAGGATTTTTC	297
Df	TTTCATTTGATTCAATGCAGTCTGGCGGTACAGCATCTCGAATCCACAGGATTTTTC	366

Db 357 TCACATCGTTGGTCACTCCATTCATCCATTGTTGTGACAAAAGTTGATCGACAT 416

Db 417 ATCGAATCAATGCTTTTACATGGAAGCTGATNNGACAGTTAGGGCTCAACTTGAAGTTT 416

Oy 410 CTGAATCAACAAGCTTTTCAAACTGAATGTCGCCGATGCCAGTATGCCGTTATGAAA 469

Oy 470 TTTTGATGCTGGACCAACCGTTCACAGTTCATTGTGCTATCATTTGTCACGCCAGTTTT
- ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 537 TTCTTGATGCTGGACCAATCCGACCAACCCTGTCATTATTTGGTACCATTTGGTCAACCAAGTCT 596

QY 530 ATCATATAATGACATGGATTCGAAACCGTTGATCTTCTGCGCGGTTGTCAT 585
 Db 597 -CCATAATGACATGATGATTCGAAACAGTTGATACATCTTGGCGTGGTGCACAT 651

RESULT 7
 BF727527 740 bp mRNA linear EST 08-JAN-2001
 LOCUS SMOV3MCA50E11SK Onchocerca volvulus molting L3 larva cDNA
 DEFINITION (SL96MLM-Ovml3) Onchocerca volvulus cDNA clone SMOV3MCA50E11 5',
 mRNA sequence.
 ACCESSION BF727527 GI:12045388
 VERSION BF727527
 KEYWORDS EST.
 SOURCE Onchocerca volvulus.
 ORGANISM Onchocerca volvulus
 Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
 Onchocercidae; Onchocerca.

REFERENCE 1 (bases 1 to 740)
 AUTHORS Williams, S.A., Lizotte-Waniewski, M., Laney, S. and Lustigman, S.
 TITLE Genes expressed in molting L3 larvae of Onchocerca volvulus
 JOURNAL Unpublished (1997)
 COMMENT Contact: Steven A. Williams
 Molecular Parasitology
 Smith College Department of Biological Sciences
 Department of Biological Sciences, Clark Science Center, Smith
 College, Northampton, MA, 01063, USA
 Tel: 4135853826
 Fax: 4135853786
 Email: genome@smith.edu
 Seq primer: Bluescript SK.

FEATURES
 source
 1..740
 Location/Qualifiers
 /organism="Onchocerca volvulus"
 /strain="Kumba, Cameroons"
 /db_xref="taxon:6282"
 /clone="SMOV3MCA50E11"
 /clone_lib="Onchocerca volvulus molting L3 larva cDNA
 (SL96MLM-Ovml3)"
 /dev_stage="molting L3"
 /lab_host="X11-Blue MRF"
 /note="Vector: Lambda Uni-ZAP XR; Site.1: Eco RI; Site.2:
 Xho I; Filarial nematode parasite of humans. Third-stage
 larvae, L3, were isolated from infected black flies in
 Cameroon (forest strain). The L3 were cultured in 20% FCS
 in IMDM+NCIC 135 and collected after day 1, 2, or 3 in
 culture. L3 of O. volvulus molt to fourth-stage larvae by
 day 5 in culture. RNA was isolated from approximately
 6000 molting larvae (ML3)/2000 larvae from day 1, 2 or 3
 in culture, and converted to double-stranded cDNA using
 reverse transcriptase and oligo(dT) followed by RNase H
 and DNA pol I. The library was constructed in the lambda
 Uni-ZAP XR vector and has 1 x 10E6 independent
 recombinants and the average insert size is ~1200 bp. The
 library was constructed by Sara Lustigman and Michelle
 Lizotte-Waniewski in the laboratory of Dr. S. A. Williams.
 The library is available from Dr. Sara Lustigman (email:
 slustig@nphc.org)."

BASE COUNT 217 a 140 c 157 g 226 t

ORIGIN

Query Match 33.4%; Score 387.2; DB 12; Length 740;
 Best Local Similarity 82.0%; Pred. No. 41e-97;
 Matches 482; Conservative 0; Mismatches 103; Indels 3; Gaps 3;

QY 50 ATTCGATTCGCGTGTGCGAGTGCAGCAATGGAATGGACCACTT 109
 Db 147 ATGCAATTCGCGTGTGCGAGTGCAGCAATGGAATGGACCACTT 206

QY 110 CATATACATCAATTTTAAATACAGTATGATTCGGAAGACATGTTATGTGAAGGTC 169
 Db 207 CAATTAAGTCAACATTTAATACCTGTAATCATTTGAAGACATGATATCGTAAGGCT 266

QY 170 TTATGATCAAGAAGTTGCCGTAATGANGAAGGTGACGTCAAGTTGCCGGAATTCAC 229
 Db 267 TATACGATGACGAGATGCCGAATGATGAGGTGACGTGACGGAATTCAC 326

QY 230 TTCCATTGATTCATGCAATGTTGCCGCTACAGATCTCGAATCCAGCGGATTTTGG 289
 Db 327 TTCCGTTGATTCCTGTAAGTAGACAGTACAGCTTGTAATCCAGCGGATTTTGG 386

QY 290 TAACCAACATCTGTTGATTCATTCCTTCATCCATTTATTTGTTACCAAGTGTGATGAT 349
 Db 387 TCATCATCAGTTGTGTCATCTCATTCATCCATTTGTTGACAAAAGTTGATGACGAT 446

QY 350 ATGAGTACATGCTTTTACATGAGGAGCAATTAACAGTACTAGTGCACAGATTGAGTAT 409
 Db 447 ATCAATACATGCTTTTACATGAGGAGCAATTAACAGTACTAGTGCACAGATTGAGTAT 506

QY 410 CTGAATTCACATGCTTTTCAACATCAATGCTCCGAGTCCAGTATCCGTTANGAAN 469
 Db 507 CCGAAATGACACATGATTTTGCAACACAAATGTATACCGATGCTGTATCCGATATGAGA 566

QY 470 TTTTGATGTGGACCAACCGGTCACACAGTTCATTTGCTATCATTTGTCAGCCAGTTT 529
 Db 567 TTCTTATGATGTGACCATCCGAC -ACGTTCAATTTGCTACATTTGTCACACAGT 625

QY 530 ATCATATAATGACATGCAATTCGAAACCGTTGATCTTCTGCGCGGTTGTCATTCCT 589
 Db 626 TCATATAATGACATGATGATTCG -AACAGTTGATTCATCTG -GGCGTGCACATTCAT 683

QY 590 GCTTGTGATGATGTAACGGTGATGACGTGGAATTTCTAATGCTG 637
 Db 684 GCTTGTGATGATGTAACGGTGATGACGTGGAATTTCTAATGCTG 731

RESULT 8
 BF400424 648 bp mRNA linear EST 28-NOV-2000
 LOCUS SMOV3MCA50E11SK Onchocerca volvulus molting L3 larva cDNA
 DEFINITION (SL96MLM-Ovml3) Onchocerca volvulus cDNA clone SMOV3MCA50E11 5',
 mRNA sequence.
 ACCESSION BF400424 GI:11386468
 VERSION BF400424
 KEYWORDS EST.
 SOURCE Onchocerca volvulus.
 ORGANISM Onchocerca volvulus
 Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
 Onchocercidae; Onchocerca.

REFERENCE 1 (bases 1 to 648)
 AUTHORS Williams, S.A., Lizotte-Waniewski, M., Laney, S. and Lustigman, S.
 TITLE Genes expressed in molting L3 larvae of Onchocerca volvulus
 JOURNAL Unpublished (1997)
 COMMENT Contact: Steven A. Williams
 Molecular Parasitology
 Smith College Department of Biological Sciences
 Department of Biological Sciences, Clark Science Center, Smith
 College, Northampton, MA, 01063, USA
 Tel: 4135853826
 Fax: 4135853786
 Email: genome@smith.edu
 Seq primer: Bluescript SK.

FEATURES
 source
 1..648
 Location/Qualifiers
 /organism="Onchocerca volvulus"
 /strain="Kumba, Cameroons"
 /db_xref="taxon:6282"
 /clone="SMOV3MCA50E11"
 /clone_lib="Onchocerca volvulus molting L3 larva cDNA
 (SL96MLM-Ovml3)"
 /dev_stage="molting L3"
 /lab_host="X11-Blue MRF"
 /note="Vector: Lambda Uni-ZAP XR; Site.1: Eco RI; Site.2:
 Xho I; Filarial nematode parasite of humans. Third-stage
 larvae, L3, were isolated from infected black flies in
 Cameroon (forest strain). The L3 were cultured in 20% FCS

EST.
pig roundworm.
Ascaris suum.
Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea;
Ascarididae; Ascaris.
1 (bases 1 to 646)
McCarteer, J., Clifton, S., Chiappelli, B., Pape, D., Martin, J., Wylie, T.,
Dante, M., Marr, A. M., Hillier, L., Kucaba, T., Theising, B., Bowers, J.,
Gibbons, M., Ritter, E., Bennett, J., Franklin, C., Tsagarishevili, R.,
Ronko, I., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe,
M., Allen, M., Pearson, B., Sallier, T., Harvey, N., Schurk, R., Kohn, S.,
Shin, Y., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and

TITLE The Washington Univ. Nematode EST Project, 1999
JOURNAL Unpublished (1999)

The Washington Univ. Nematode EST Project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Fax: 314 286 1810
Email: est@watson.wustl.edu
The library was supplied by Dr. Dante Zarlenga of the USDA Immunology and Disease Resistance Lab in Beltsville, MD. The cDNAs were made from 21 day I4 proximal (isolated from jejunum) and directionally cloned into the NotI/SalI sites of pSPORT1. Seq primer: -40RP from Gibco. High quality sequence stop: 482.

FEATURES	Location/Qualifiers
source	1. 646

```

/organism="Ascaris suum"
/db_xref="taxon:6253"
/clone_lib="Ascaris suum 14 nspDM1 Zarlenka v1"

```

```
/usr/bin/nc -l -p 4444 -e /bin/bash
/dev_stage="L4"
/lab_host="DH5-alpha"
```

```

Vector: pSPOR1 (Life Technologies); Site_1: NotI;
Site_2: SalI; The library was supplied by Dr. Dante

```

Zarlenga of the USDA Immunology and Disease Resistance Lab in Beltsville, MD. The cDNAs were made from 21 day L4

proximal (isolated from jejunum) and directionally cloned into the NotI/SalI sites of pSPORT1."

BASE COUNT	132 d	173 c	108 g	132 l	1 oliveis
ORIGIN					

Query Match	32.5%	Score 376.8;	DB 13;	Length 646;
Best local Similarity	74.4%	Pred. No. 3 2e-94;		

Matches 474; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

QY 32 CACATATTGCATTGCTTATTTCGATTCCGGTTTGACAATAGTGTGCGAAGGTGAGGCCAGAAA 91
| |||| | | |||||| | ||||| | ||||||| ||||| ||| -
QZ

D6 CTCTTTGGGCTCGCCGACGATTCCTGTGACACGAGAGTCGAGGAGAGCCGGAGA 65

92 TTGAATGTGCACCAACATTCATTAACATCAATTTTAAACAGCTAAAGCATTCGAGAGAC 151

	08	116AA1G1B61CC1ACG1CAH1CACGG1CAAC11CAAAC1CAGAGAAL1CA111CGAGGG1C	125
DU			
YV	152	AACCTTATATTCCTGAAGAACCTTGCTTTATATGATACAGAACCTTGTCTTCTTAATGCATGAAACCTTGAGCTTCTC	211

126 ATGTATACGTAAGGACCTTATGATCAAGAGGGATCCGACCGACAGAGGGGTGGACCTC 185

212 AAGTTGCCGGAATTTTCACCTTCATTTGATTCAATGCATGTTGGCGGTACACGATCTCTGA 271

186 AAGTGGCTGGCATCTTCGCTTTTTCACTCGTGCAACGTCGACGAAACACGTCGCTGA 245

272 ATCCAGTGTATTTTGTACACAACACTGTTGTCATTTGTTCAACCATTTATTTGTTA 331

Db 246 ATCCCTGGTAICTTTTGTCACGACACGGTTGTCATCTCGTTCCATCCGCTTTCATCA 305

332 CCAAGTGTATCGTCATATCGAGTACATGCTTTTACATGCAAGCTGATAAACAAGTTA 391

Db 306 CCAAAGTTGACCCGCTACCGCGCTCCATGCTTCTACATGGAAGCCGACAAAGACTGTCA 365

Q7 392 GTGCACGATTGAGGATATCTGAAATCACAACACTGCTTTTCAAACTCAATTTGCCGATGC 451

Db 366 GTACGCAATGCAAGTGTCTGAGATTGACACCGCTTCCAGACCAATGCTCCATGTC 425
QY 452 CAGTATGCCCTTATGAAATTTGGATGGTACCAACCGGTCAACAGTCAATTTGCTA 511
Db 426 CAGTCTGCAATGAGATTCCTGACGAGAGTCCACAGGAGCCGATACAGTGGCAA 465
QY 512 TCATGTGTCACCCAGTTATATCAATATGACATGCGATTTCTGAACCGTTGATCTTCT 571
Db 486 CCATCGGTGACGAGTATACACAGGAGCACTTGCATTTGGAACCTGTGATACGTTCT 545
QY 572 GCGCGGTTCCTCTCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 631
Db 546 GTCTGTGCTTCACTACTGCTGTGATGATGATGATGATGATGATGATGATGATGATGAT 605
QY 632 ATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 668
Db 606 ACGAGGAGGAGTGTGCGCTTGACAAATACCTGCTGAA 642

RESULT 10
Bg809012 747 bp mRNA linear EST 22-MAY-2001
LOCUS SMOV3MCA1144SK Onchocerca volvulus molting L3 larva cDNA
DEFINITION (SL96MLM-OvML3) Onchocerca volvulus cDNA clone SMOV3MCA1144 5',
mRNA sequence.
ACCESSION Bg809012 GI:14180004
VERSION Bg809012.1 GI:14180004
KEYWORDS EST.
SOURCE Onchocerca volvulus.
ORGANISM Onchocerca volvulus.
REFERENCE 1 (bases 1 to 747)
AUTHORS Williams, S.A., Lizotte-Waniewski, M., Laney, S. and Lustigman, S.
TITLE Genes expressed in molting L3 larvae of Onchocerca volvulus
JOURNAL Unpublished (1997)
COMMENT Contact: Steven A. Williams
Molecular Parasitology
Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith
College, Northampton, MA, 01063, USA
Tel: 4135853826
Fax: 4135853786
Email: genomesmith.edu
Seq primer: pluscript SK.
Location/Qualifiers
1..747
/organism="Onchocerca volvulus"
/strain="Kumba, Camerouns"
/db_xref="taxon:6282"
/clone="SMOV3MCA1144SK"
/clone_lib="Onchocerca volvulus molting L3 larva cDNA
(SL96MLM-OvML3)"
/dev_stage="molting L3"
/lab_host="XLI-Blue MRF"
/note="Vector: Lambda Uni-ZAP XR; Site.1: Eco RI; Site.2:
Xho I; Filarial nematode parasite of humans. Third-stage
larvae, L3, were isolated from infected black flies in
Cameroun (forest strain). The L3 were cultured in 20% FCS
in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in
culture. L3 of O. volvulus molt to fourth-stage larvae by
day 5 in culture. mRNA was isolated from approximately
6000 molting larvae (ML3), 2000 larvae from day 1, 2 or 3
in culture, and converted to double-stranded cDNA using
reverse transcriptase and oligo(dT) followed by RNase H
and DNA pol I. The library was constructed in the lambda
Uni-Zap XR vector and has 1 x 10⁶ independent
recombinants and the average insert size is ~1200 bp. The
library was constructed by Sara Lustigman and Michelle
Lizotte-Waniewski in the laboratory of Dr. S. A. Williams.
The library is available from Dr. Sara Lustigman (email:
slustigman@nc.org)."

BASE COUNT 210 a 137 c 159 g 241 t
ORIGIN
Query Match 32.4%; Score 376.2; DB 12; Length 747;
Best Local Similarity 80.4%; Pred No. 4.9e-94;
Matches 514; Conservative 0; Mismatches 118; Indels 7; Gaps 6;
QY 50 ATTGATTCGCGTGGCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 109
Db 109 ATGCTATTCGCGTGGTAAACGGGTGTAAGGATGATGATGATGATGATGATGATGAT 168
QY 110 CAATTAACATCAATTTTAAATACAGTAATGATGATGATGATGATGATGATGATGAT 169
Db 169 CAATTAACATCAATTTTAAATACAGTAATGATGATGATGATGATGATGATGATGAT 228
QY 170 TTTATATCAAGAGGTTGCGGTAATGATGATGATGATGATGATGATGATGATGATGAT 229
Db 229 TATACATCAAGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 288
QY 230 TTGCAATTTGATTCATGCAATGATGATGATGATGATGATGATGATGATGATGATGAT 289
Db 289 TTGCAATTTGATTCATGCAATGATGATGATGATGATGATGATGATGATGATGATGAT 348
QY 290 TACAAACATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 349
Db 349 TCACATCAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 408
QY 350 ATGCAATCAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 409
Db 409 ATGCAATCAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 468
QY 410 -CTGAATACCAACTGC-CTTTCAACTCAATTTGCTCCGATGCCAGTATGCCGTTATGA 467
Db 469 CCCGAATGATCAACTGCCTTTGCAACTCAATTTGCTCCGATGCCAGTATGCCGTTATGA 528
QY 468 AATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 527
Db 529 GATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 588
QY 528 TTATCAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 586
Db 589 TTGCAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 648
QY 587 CCTGCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 644
Db 649 CATGCTTCGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 708
QY 645 TGCTCTT--GATTAATATTTGCTTAATATTTGGAATAT 681
Db 709 GGGCTTTGGACAAATATTTGCTCAACACCTTGATAT 747

RESULT 11
AA701731 628 bp mRNA linear EST 19-DEC-1997
LOCUS SMOV3MCA1144SK Onchocerca volvulus molting L3 larva cDNA
DEFINITION (SL96MLM-OvML3) Onchocerca volvulus cDNA clone SMOV3MCA1144 5',
mRNA sequence.
ACCESSION AA701731 GI:2704931
VERSION AA701731.1 GI:2704931
KEYWORDS EST.
SOURCE Onchocerca volvulus.
ORGANISM Onchocerca volvulus.
REFERENCE 1 (bases 1 to 628)
AUTHORS Williams, S.A., Lizotte-Waniewski, M., Laney, S. and Lustigman, S.
TITLE Genes expressed in molting L3 larvae of Onchocerca volvulus
JOURNAL Unpublished (1997)
COMMENT Contact: Steven A. Williams
Molecular Parasitology
Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith


```

/organism="Onchocerca volvulus"
/strain="Kumba, Cameroons"
/db_xref="taxon:6282"
/clone="SMOVC3MCAM49C08"
/clone_id="Onchocerca volvulus molting L3 larva cDNA
(SJ96MW-Ovml3)"
/dev_stage="molting L3"
/lab_host="XLI-Blue MRF"
/notes="Vector: lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:
Xho I; Filarial nematode parasite of humans. Third-stage

```

QY	50	ATTGATGATCCCGGTGACATAGTGTGCGAAGGTGAGCCCGAAATTTGAATGAGCAACACTT	109
Db	170	ATGCAATTCGGGTGATATACGGTGTAGAAAGGTGAGCAACGAATTTGAATGTGGTCCAAACAT	229
QY	110	CAATTAACATCAATTTTAAATACAGCTAAATGCAATTCGGAAGACATGTTTATGTGAAGGTC	169
Db	230	CAATTAAGTGTCAACTTTTAATCTGTAATTCATTTGAAGAGACATGTATACGTGAAGAGCT	289
QY	170	TTTATGATCAAGAAAGTTCCCGTAAATGATGAAGGTGACGCTCAAGTTCGCCGAATTTTAC	229
Db	290	TATAGATCCAGGAGGATGCCGAATGTATGAAGGTGACGCTCAGGTACCGGAATTTGAC	349
QY	230	TTTCATTTGATTCATGCAATGTGGCGCTAACACATCTCGAATCCAGTGGTATTTTG	289
Db	350	TTCCGTTGATTTCCGTATATGTATGACGCTACACGTTCCGTAAATCCAGCTGGTATTTTG	409
QY	290	TAAACAACAACGTTGTGTCATTTCTGTTTCATCCATTTATTTGTAACAAAGTTGATCGTGAT	349
Db	410	TCACATCAGTGTGTGTCATCTCATTCATTCATTTGTTGTGACAAAGTTGATGACGAT	469
QY	350	ATCGAGTAGCAATGCTTTTACATGGAAGCTGATATAAACAGATTAGTGCACAGATTGAGTAT	409
Db	470	ATCGAATTCATGCTTTTACATGGAAGCTGATATGAACAGTTAGCGCTCACACTTGAAGTTT	529
QY	410	-CTGAATCACAACTGCTTTTTCAAACTCAAAATTGTCCGATGCCAGTAGTGCCTGATTATGAA	468
Db	530	CCCCAAATGACAACTGCACTTTTGCAACACAAATTTG-CGATGCGCTGTATGCCGATATGAG	588
QY	469	ATTTTGGATGGTGGACCAACCGGTCAACC-AGTTCAAATTTGCTATCATTTGGTCAAGCCAGT	527
Db	589	ATTTTGAATGGTGGACCATTCGGACACCTGTTGCAATTTGCTACCATTTGGTCAACAGT	648
QY	528	TTATCATTAATGAGCATGCGCATCTGGAACCGCTTG--ATACTTCTGGCGGTTGTCAT	585
Db	649	GGTCCATTAATGACATGTGATTTCTGANAACAGTTGGATACCATTTCTGTGCTGGTGACCAT	708
QY	586	TCCCTCTTTGTGATGATG--TAAAGGTGATCTGGAATTTCTAAATCTGATGAT	643
Db	709	TCATGCTTCCTGATGATGGTAAACCGGTGATTAAGGGGAGATTTTAAATTTGCTGTGAT	768
QY	644	GTGCT 648	
Db	769	GGGTT 773	

RESULT 14

LOCUS

DEFINITION

AA618895

AA618895

672 bp

SMOVC/MCA1879SK Onchocerca volvulus molting L3 larva cDNA (SL96MLM-OvML3) Onchocerca volvulus cDNA clone SMI13CO1879 5', mRNA sequence.

MRNA

linear

EST 12-NOV-1997

Accession	Version	Keywords	Source	Organism
AA618895	AA618895.1	GI:2522771	EST	Onchocerca volvulus.
				Onchocerca volvulus.
				Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
				Onchocercidae; Onchocerca.
REFERENCE	1 (bases 1 to 672)			
AUTHORS	Williams, S.A., Lizotte-Waniewski, M., Laney, S. and Lustigman, S.			
TITLE	Genes expressed in molting L3 larvae of Onchocerca volvulus			
JOURNAL	Unpublished (1997)			
COMMENT	Contact: Steven A. Williams Molecular Parasitology Smith College Department of Biological Sciences Department of Biological Sciences, Clark Science Center, Smith College, Northampton, MA, 01063, USA Tel: 4135853826 Fax: 4135853786 Email: genome@smith.edu Seq primer: Bluescript SK. Location/Qualifiers			
FEATURES	1..672			
source	/organism="Onchocerca volvulus"			
	/strain="Kumba, Camerouns"			
	/db_xref="taxon:6282"			
	/clone="SMnl3CO1879"			
	/clone_lib="Onchocerca volvulus molting L3 larva cDNA (SL56MW-OvML3)"			
	/dev_stage="molting L3"			
	/lab_host="XLI-Blue MRP"			
	/note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2: Xho I; Filarial nematode parasite of humans. Third-stage larvae, L3, were isolated from infected black flies in Cameroon (forest strain). The L3 were cultured in 20% FCS in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in culture. L3 of O. volvulus molt to fourth-stage larvae by day 5 in culture. mRNA was isolated from approximately 6000 molting larvae (ML3), 2000 larvae from day 1, 2 or 3 in culture, and converted to double-stranded cDNA using reverse transcriptase and oligo(dT) followed by RNase H and DNA pol I. The library was constructed in the lambda Uni-Zap XR vector and has 1 x 10 ⁶ independent recombinants and the average insert size is ~1200 bp. The library was constructed by Sara Lustigman and Michelle Lizotte-Waniewski in the laboratory of Dr. S. A. Williams. The library is available from Dr. Sara Lustigman (email: slustigman@bc.org)."			
BASE COUNT	198 a 132 c 129 g 202 t			11 others
ORIGIN				
Query Match.	31.4%;	Score 364.2;	DB 9;	Length 672;
Best Local Similarity	81.1%;	Pred. No. 1,1e-90;		
Matches	417;	Conservative 0;	Mismatches 97;	Indels 0;
			Gaps 0;	
50	ATTCGATTCGGTTGACATGTTGTGCGAGGTGAGCCAGAAATTTGAATGTGACCAACTT	109		
159	ATGCTATTCGGGTGATAGCGGTGATGAGAGGTGAGCAACCAATTTGAATGTGTCACAACT	218		
110	CAATACATCAATTTTATATACAGCTAATGCAATTTGCAAGCAATGTTTATGTGAAAGTGC	169		
219	CAATTACGTCGCAACTTATATCTGTAATCCATTTTGAAGCAATGATATACGTTGAAGGCT	278		
170	TTTATGATCAACAAGTTCGGTAAATGATGAGGTGAGCGTCAAGTTGGCCGAATTTGAC	229		
279	TATAGCATGAGCAGAGATATCCGAATGATGAAGGTGACGTACAGTACGCCGCAATTTGAAC	338		
230	TTTCATTTGATTCATGCAATGTTGGCGGTACACAGATCTCTGCAATCCACGTGATTTTGG	289		
339	TTCCGTTGATTCGTGATATGATGACGCTACGCTTCGTTAATCATCCAGTGGATTTTGG	398		
290	TACACACACACTGTGTCATTTGCTTTCATTCATTAATTTTTCACAAAGTTGATCGCAT	349		
399	TCACTACAGTTTGTCCNNCTCATTCATCAATGTTGTTGTGACAAAGTTGATCGCAT	458		

Query Match	Best Local Similarity	Score	DB	Length	Matches	Conservative	Pred.	Mismatches	Indels	Gaps													
330	ATCGAGTACATGCTTTTACATGGAAGCTGATATAAACAGTTATGTCACAGATGTAGGAT	409	459	ATTCGATTAACAAAGCTTTTACATGGAAGCTGATATAAACAGTTATGTCACAGTGAATTT	518	410	CTGAATACCAACGCTTTTCAACTGCAAAATGTGCCGATGTCAGTATGCCGTTATGAA	469	519	CCGAATATGACACTGCTATTGTCNACACAAATTTGTACCGATGCTTATATCCCATATGAGA	578	470	TTTGGATGCTGGGACCAACCGCTTACACAGTTTCATTTGCTATCATTTGGTCAGCAGATT	529	579	TTCTTGATGGTGGACCATTCGCGACAAACCTGTTCACTTTGCTATNCATPTGGTCMACCAGTGT	638	530	ATCATTAATGACATCGCATTTCTGAAACCCGTTGA	563	639	TCCATATNATNGAANATNTGATTCGTGAACAAGTTGA	672

```

OY 41 CATTGCTTATTTGGATTCGGGTGACAAATGCTGTCCAGAGTGAACCAAAATTGATGTG 100
    ||||| || ||||| ||||| || ||||| || ||||| || ||||| || ||||| ||
Db 15 CATTACTGTGGGATCCCGGTGATTAACGGCTTGGAGGAGAACCGGAATTGAATGTG 74
OY 101 GACCACTTCATACATCAATTTTAAATACGATATGATGATGCAAGCATTTTATG 160
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 75 GCGCCACCTCAATTAACATCACTTTTACCTGAAATTCATTTGAGGGGACGTTTACG 134
OY 161 TGAAGGCTTTTATGATCAAGAGGTTGCCGTAATGATGAGGTGACGTCAAGTTGCCG 220
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 135 TAAAGACCTATGATCAAGAGGATGCCGTAACGATGATGATGAGCGTAGGTTGCCG 194
OY 221 GAATTCACCTTCATTTGATTCATGCAATGTTCCGCTACACATCTGCAATCCAGTG 280
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 195 GATTCACGTTACCGTTGCAATGTTGCAAGATGTTGCAAGAACCCCTCACTTAAATCCAGCG 254
OY 281 GATTTTGTAAACAAACAGCTGTTCATTCATTCATTCATTTGTTTACCAAGTTG 340
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 255 GCATCTTCGTGACGACGACTGTGTATCTCATTTTCACTCTTCCTACCAAAAGTGC 314
OY 341 ATCGTCATATGAGTACATGCTTTTACATGAGACGTGATTAACAGATTAGTGCACAG 400
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 315 ATCGACCTATGCGCATTCAGTGTCTTACATGGAACCGCAGACAGATCAGACATCAGC 374
OY 401 TTGAGTATCTGAATACACAACTGCTTTTCAAACTGCAAAATGTCCTCCGATGCCAGTATGCC 460
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 375 TAGAATGTGCGAGATTAACGACTGCTTCACACTCAGATTTGTTCCGATGCCAGTTGCA 434
OY 461 GTTATGAATTTTGGATGTTGGACCAACGGGTCAACCACTTCAATTTGCTATCATTTGTC 520
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 435 GATACAGATTTCTCGATGTGTGACCTAGTGACACCAATTAATTTGCTATTAATTTGGTC 494
OY 521 AGCCAGTTTATCAATAATGACATGCGATTTGAAACCGTTGATTTCTGCGCGGTTG 580
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 495 AACAGATATATCAAGAGTGTGACTGTGACCTCGGAAACTTTGACACATTTCTGTGCGATTTG 554
OY 581 TCCATTCCTGCTTTGCTGATGATGTAACGGTGTATCTGTGAAATTTCTTAATG 634
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 555 TCCATTCAGTTTCTGTCAGATGTAATGTGACAAAGTGAACATCATCAATG 608

RESULT 16
AI322117 639 bp mRNA linear EST 22-DEC-1998
LOCUS AI322117
DEFINITION SMOV3MCAM12G08SK Onchocerca volvulus molting L3 larva cDNA
(SU96MLM-Ovml3) Onchocerca volvulus cDNA clone SMOV3MCAM12G08 5',
mRNA sequence.
ACCESSION AI322117
VERSION AI322117
KEYWORDS EST.
SOURCE AI322117.1 GI:4056268
ORGANISM Onchocerca volvulus.
Onchocerca volvulus.
Eukaryota; Metazoa; Nematoda; Chromadorea; Sphurida; Filarioidea;
Onchocercidae; Onchocerca.
REFERENCE 1 (bases 1 to 639)
AUTHORS Williams,S.A., Lizotte-Waniewski,M., Laney,S. and Lustigman,S.
TITLE Genes expressed in molting L3 larvae of Onchocerca volvulus
JOURNAL Unpublished (1997)
COMMENT Contact: Steven A. Williams
Molecular Parasitology
Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith
College, Northampton, MA, 01063, USA
Tel: 4135853826
Fax: 4135853786
Email: genome@smith.edu
Seq primer: pBluescript SK.
FEATURES
Source
1..639
Location/Qualifiers
/organism="Onchocerca volvulus"
/strain="Kumba, Cameroons"
/db_xref="taxon:6282"

```

```

/clone="SMOV3MCAM12G08"
/clone_lib="Onchocerca volvulus molting L3 larva cDNA
(SU96MLM-Ovml3)"
/dev_stage="molting L3"
/lab_host="XLI-Blue MRF"
/note="Vector: Lambda Uni-Zap XR. Site.1: Eco RI; Site.2:
Xho I; Filarial nematode parasite of humans. Third-stage
larvae, L3, were isolated from infected black flies in
Cameroon (forest strain). The L3 were cultured in 20% FCS
in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in
culture. L3 of O. volvulus molt to fourth-stage larvae by
day 5 in culture. mRNA was isolated from approximately
6000 molting larvae (ml3), 2000 larvae from day 1, 2 or 3
in culture, and converted to double-stranded cDNA using
reverse transcriptase and oligo(dT) followed by RNase H
and DNA pol I. The library was constructed in the lambda
Uni-Zap XR vector and has 1 x 10E6 independent
recombinants and the average insert size is ~1200 bp. The
library was constructed by Sara Lustigman and Michelle
Lizotte-Waniewski in the laboratory of Dr. S. A. Williams.
The library is available from Dr. Sara Lustigman (email:
slustig@nbc.org)."
BASE COUNT 197 a 130 c 121 g 191 t
ORIGIN
Query Match 30.9%; Score 358.6; DB 9; Length 639;
Best Local Similarity 84.5%; Pred. No. 3.8e-89;
Matches 403; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

OY 50 ATTCGATTCGGGTGACAAATGCTGTGCAAGGTGAGCAGAAATTGAATGACCAACTT 109
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 162 ATGCTATTCGGGTGATACGGTGTAGAGTGAACCAAAATTGAATGATGCTCAACT 221
OY 110 CATATACATCAATTTTATACACGTAATGATTCGAAGCAATGTTATGGAAGTTC 169
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 222 CATTAATGCTCAACTTAATTAATCTGTAATCACTTTGAAGCAATGTAATGGAAGCT 281
OY 170 TTTATGATCAAGAGGTGCGTAATGATGAGGTGAGCAAGTTCGGAATTTTCAC 229
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 282 TATGAGATGAGCAGAGATGCCAANTGATGAGGTGAGTACAGTATGACCGGAATTTGAC 341
OY 230 TTCCATTTGATTCATGCAATGTTGCGCGTACACGATCTCTGATCCAGCTGATTTTGG 289
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 342 TTCCGTTGATTCCTGTAATGTAGACACGTTCGTTAATTCACGCTGATTTTGG 401
OY 290 TAAACAACACTGTGTCATTTGCTTTCATTCATTTTGTTCGCAAGTGTATGCGCAT 349
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 402 TCACATCACTGTTGTGCTATCTATTCATTCATCTTTGTTGTGCAAAAGTTGATCGACAT 461
OY 350 ATCGAGTACAAATGCTTTTACATGAGAGCTGATTAACAGTATAGTCCAGATTTGAGTAT 409
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 462 ATCGAATCAAAATGCTTTTACATGAGAGCTGATTAACAGTATAGTCCAGATTTGAGTAT 521
OY 410 CTGAATACAAATGCTTTTCAACTCAATATGTTCCGATGCGCATATGCCGTTATGAAA 469
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 522 CCGAATGACAACTGCTATTTGCAACAAATTTGATCCGATGCTGATGCGATTTGAGA 581
OY 470 TTTTGGATGTGAGCAACCGTCAACAGTTCAATTTGCTATCAATTTGCTGACGAGCAG 526
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 582 TTCTTGATGTGAGCAATCCGACACACTGTTCAATTTGCTACACTTTGCTACACCG 638

RESULT 17
AA668071 610 bp mRNA linear EST 20-NOV-1997
LOCUS AA668071
DEFINITION SMOV3MCAM02C01SK Onchocerca volvulus molting L3 larva cDNA
(SU96MLM-Ovml3) Onchocerca volvulus cDNA clone SMOV3MCAM02C01 5',
mRNA sequence.
ACCESSION AA668071
VERSION AA668071
KEYWORDS EST.
SOURCE Onchocerca volvulus.
ORGANISM Onchocerca volvulus

```

Db 302 CGATACGAGATTCTGAGAGGTGGACCAACACGTTACACCTATTCGATATGCACATGATCGGA 361

BASE COUNT	157 a	150 c	171 g	168 t	2 others
ORIGIN					

library was constructed by Sara Lustigman and Michelle Lizotte-Waniewski in the Laboratory of Dr. S. A. Williams. The library is available from Dr. Sara Lustigman (email: slustig@u.washington.edu).

BASE COUNT	156 a	106 c	112 g	180 t
ORIGIN				

Query Match	28.5%	Score 331.4	DB 12	Length 554
Best Local Similarity	84.2%	Pred. No. 1.4e-81		
Matches 385	Conservative 0	Mismatches 71	Indels 1	Gaps 1

QY	50	ATTGATGATCCGGTTGACAAATGCTTCGAAGGTGAGCCAAATGAAATGGAGCAACTT	109
Db	94	ATGCTATTCCGGTTGATPACGGTGTAGAAAGTGAACCAAAATTGAATTGTGTCCAAAT	153
QY	110	CAATAACAAATCAATTTTAATACACGTAATGCATTGCAAGGACATGTTATGTGAAGTGC	165
Db	154	CAATTACTGTGCACCTTAATTAATCTGTGAATCCATTGGAAGGACATGTAACTGAAAGGCT	213
QY	170	TTTATGATCAAGAAGGTTCGGCTAATGATGAAGTGGACGTCAGTTCGGGAATTTTCAC	222
Db	214	TATACGATCAGGCGAGGATGCGGAATATGAAGGTGACGTCAGGTGACGGGAATTTAAC	273
QY	230	TTTCATTGATTCATGCAAAATGTTCCGGGTACACAGCATCTGTAATCCACAGTGTATTTTG	289
Db	274	TTCCGTTTGATTCCTGTAAATGTAGACAGTACAGCTTGTTAATCCACGTGTATTTTTCG	333
QY	290	TACACAAACTGTGTGCAATTCGTTTCATCCATTAATTTGTTACCAAGTTGATGTCAT	349
Db	334	TCACTATGATGTTGTGTCATTCATCCATCCATGTTTGTGACAAAGTTGATGACAT	393
QY	350	ATCAGTACATGCTTTTACATGGAACCTGATAAACAAGTTAGTGCACAGATTGAGGTAT	409
Db	394	ATCGAATACATGCTTTTACATGGAACCTGATAAACAAGTTAGTGCCTCAACTGGAATTT	453
QY	410	CTGAATACAACTGCTTTTCAAACTCAAAATGTCCGATGCCAGTATGCCGTTATTTAAA	463
Db	454	CCGAAATAGACAACGTCAATTGGCAACACAAATTGT-CCGATGCTGTATGCCGATATAGGA	512
QY	470	TTTTGATGTGTGACCAACCGGTCAACACATGTCATTT	506
Db	513	TTCTTGTATGTGTGACCAATCCGGGAACCTGTTCATTT	549

RESULT 22	LOCUS	776 bp	mRNA	linear	EST 30-DEC-1997
AA585626	AA585626				
DEFINITION	SM3D9CA349SK	Brugia malayi L3 molting-day 9 larva	CDNA		
	(SAM97MLM-BML3d5)	Brugia malayi	CDNA clone	SM3D9CA349	5', mRNA sequence.

/org_xrefsm="Brugia malayi"
 /db_xrefsm="taxon:6279"
 /clone="SM3DCA349"
 /clone_id="Brugia malayi L3 molting-day 9 larva cDNA
 (SAW9/MLW-Bml3d9)"
 /dev_stage="third stage larvae, nine days after infection"
 /lab_host="E. coli XL1-Blue MRF"
 /note="Vector: lambdaZap II (UnizAP XR); Site_1: Eco RI
 (5' end); Site_2: Xho I (3' end); Brugia malayi is a
 lymphatic filarial nematode parasite of humans. mRNA was
 prepared from third stage larvae of Brugia malayi
 isolated from the peritoneal cavity of birds nine days
 after infection. The mRNA was converted to double
 stranded cDNA using reverse transcriptase and oligo (dT)
 followed by Rnase H and DnApol I. The library was
 constructed by Michelle Litovle-Wanlewski. The library is
 available from the Filarial Genome Project Resource
 Center: contact Dr. S.A. Williams, Clark Science Center,
 Smith College, Northampton, MA 01063 USA phone +1 413
 585-3826 fax +1 413 585-3786 email genom@smith.edu

Query Match	27.9%	Score 324.2	DB 9	Length 776
Best Local Similarity	85.2%	Pred. NO. 1.7e-79		
Matches 362; Conservative	0	Mismatches 63	Indels 0	Gaps 0

OY	737	CACAGCTTTTCATCAATGCCAGATCAAGTATTACCATTTAAAGAACCAAAATTCGCAATGTG	796
Db	1	CACAACTTTTTCATCAATGCCAGATCAAGTATTACCATTTAAAGAACCAAAAGTCAATGTG	60
OY	797	TTGCAACCAAAATGTTTCAGAACCCAGAGATTCGGAGCTGTTTAAACAGGTGTGCCCCGAG	856
Db	61	CTCGACCCACAGTGTTCAGAGCCGACAGAGATTTGGAGCTGTAAAAACGGCCGCTCCCGG	120
OY	857	CAAAACCGTGTGACGAGTGGCCGACACTGGTTTACGTCAAGAAAGATCTGCAGAACCCGAG	916
Db	121	TTTAAACCTGCAGACGAGCTGTCAACTCCGCTTGCTCAAGAAAGGTGTGTGGAAACCAAGA	180
OY	917	ATATCATGTATGATGACGAATCTGATATCAACACCCCTTGAATTAAGCATGATATATCAAGCTT	976
Db	181	ACGTGCTGACGTGAGGACCTGATATTAGCGCCCTTGAATATCAGGAGATTAATCAAGCTC	240
OY	977	TGCCAGTGGATTTCACGTACCCGTGCACCTCTTCGCAACAATATAGCAACACTGTATCTTG	1036
Db	241	TACCACTGGACTTACGTCATGCTGCACGTTTGACATCTGGTGGCAACAACCACTGGTCTCG	300
OY	1037	CTCGAGTACAAATAGCAATCTGCATGTACCATTTGGCTTCTCATGTTTATGCGTTTAA	1096
Db	301	CCGCACTGCCAAAGAGAAATTGCATGTCCACATTTTGGCTTCTCATGTTTATGGGTTTGA	360
OY	1097	GCATTTGCATTGATGTGTCGCCGTCAATATTACCAATTTGTTTTAAATTTGTCGCAAAATCAGA	1156
Db	361	GCATTTGCATTGATGTGTCGCCGTCAATGTCAACATTTCTCTGAATTTGTCGTCGAATGCAAA	420
OY	1157	AGGCA 1161	
Db	421	AAGCA 425	

JOURNAL	Brugia malayi	RESULT 23
COMMENT	Unpublished (1997)	AA668051
	Contact: Steven A. Williams	LOCUS
	Molecular Parasitology	DEFINITION
	Smith College Department of Biological Sciences	
	Department of Biological Sciences, Clark Science Center, Smith	
	College, Northampton, MA, 01063, USA	
	Tel: 4135853826	ACCESSION
	Fax: 4135853786	VERSION
	Email: genomesmith.edu	KEYWORDS
	Seq primer: pbluescript SK.	SOURCE
FEATURES	Location/Qualifiers	ORGANISM
source	1..776	

Db 202 TATACGATCAGGACAGAGCCGAAATGATGAAGTGGACGTCAGTAGCCGGAAATTTGAAC 261
 Qy 230 TTCCATTTGATTCATGATCATGTTGGCGGTACACGATCTCTGAATCCACGTGGTATTTTG 289
 Db 262 TTCCGTTTGATTCGATATAGTAGACAGTTCACGTTCTTAATCCACCTGGTATTTTG 321
 Qy 290 TAACAACAACGTGTGTGATTTGCTTTCATTCATTTATTTGTAACAAAGTTGATTCGAT 349
 Db 322 TCACATCAGTTGTGTGATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 381
 Qy 350 ATCGATGCAATGCTTTTACATGATGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 409
 Db 382 ATCGATGCAATGCTTTTACATGATGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 441
 Qy 410 CTGAATACACATGCTTTTCAACTCAATATGT-CCCGATGCCAGTATGCCGTATGAA 468
 Db 442 TCCAAATGACACATGCTTTTCAACTCAATATGT-CCCGATGCCAGTATGCCGTATGAA 501
 Qy 469 ATT-TTGATGTTGACCAACCGGTCAAC--AGTTCAATTTGCTATTCAT--GGTCAAC 523
 Db 502 ATTTCTGGATGGTGGACATGGGACAACTGTTCAATTTGGTACATTTGGTCAAC 561
 Qy 524 CAGTTATCATTAATGAGCA-TGGCATCTGAAACCGTTGATA---CTTTCGGCGGCT 579
 Db 562 AAGTATCACAATGATGACATGATGATTTGTAACAGGTTGATCAATTTGCTGCTGTT 621
 Qy 580 GTCCATTCCTGCTTGTGATGAT---GGTAAACGCTGATGCTGTAATTTCTTAA 632
 Db 622 GTACAAATTCATGCTTCCGTTGGATGATGGTAAAGGTTAAAGGCGGATTTATTA 678

RESULT 25
 A1322068 552 bp mRNA linear EST 22-DEC-1998
 LOCUS SMOV3MCA12A04SK Onchocerca volvulus molting L3 larva cDNA
 DEFINITION (SL96MLM-Ovml3) Onchocerca volvulus cDNA clone SMOV3MCA12A04 5',
 mRNA sequence.
 ACCESSION A1322068
 VERSION A1322068.1 GI:4056219
 KEYWORDS EST.
 SOURCE Onchocerca volvulus.
 ORGANISM Onchocerca volvulus.
 Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
 Onchocercidae; Onchocerca.

REFERENCE 1 (bases 1 to 552)
 AUTHORS Williams, S.A., Lizotte-Waniewski, M., Laney, S. and Lustigman, S.
 TITLE Genes expressed in molting L3 larvae of Onchocerca volvulus
 JOURNAL Unpublished (1997)
 COMMENT Contact: Steven A. Williams
 Molecular Parasitology
 Smith College Department of Biological Sciences
 Department of Biological Sciences, Clark Science Center, Smith
 College, Northampton, MA, 01063, USA
 Tel: 4135853826
 Fax: 4135853786
 Email: genome@smith.edu
 Seq primer: Bluescript SK.

FEATURES
 source
 1..552
 Location/Qualifiers

1..552
 /organism="Onchocerca volvulus"
 /strain="Kumba, Cameroons"
 /db_xref="taxon:6282"
 /clone="SMOV3MCA12A04"
 /clone_lib="Onchocerca volvulus molting L3 larva cDNA
 (SL96MLM-Ovml3)"
 /dev_stage="molting L3"
 /lab_host="X1-Blue MRF"
 /note="Vector: Lambda Uni-ZAP XR; Site.1: Eco RI; Site.2:
 Xho I; Filarial nematode parasite of humans. Third-stage
 larvae, L3, were isolated from infected black flies in
 Cameroon (forest strain). The L3 were cultured in 20% FCS
 in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in
 culture. L3 of O. volvulus molt to fourth-stage larvae by

day 5 in culture. mRNA was isolated from approximately
 6000 molting larvae (ml3), 2000 larvae from day 1, 2 or 3
 in culture, and converted to double-stranded cDNA using
 reverse transcriptase and oligo(dT) followed by RNase H
 and DNA pol I. The library was constructed in the lambda
 Uni-Zap XR vector and has 1 x 10⁶ independent
 recombinants and the average insert size is ~1200 bp. The
 library was constructed by Sara Lustigman and Michelle
 Lizotte-Waniewski in the laboratory of Dr. S. A. Williams.
 The library is available from Dr. Sara Lustigman (email:
 slustigmen@bc.org)."

BASE COUNT 154 a 106 c 111 g 181 t
 ORIGIN

Query Match 26.4%; Score 306.8; DB 9; Length 552;
 Best Local Similarity 83.4%; Pred. No. 1,1e-74;
 Matches 372; Conservative 0; Mismatches 72; Indels 2; Gaps 2;
 Qy 50 ATTCGATTCGGTTGACATGTTGTCGAGGTGACCCGAAATGATGAGCAACTT 109
 Db 105 ATGCTATTCGGTTGATAGCGTGAAGGAGAACAGAAATGATGTTGCCAACAT 164
 Qy 110 CAATACATCAATTTTATACAGTAATGCAATTTGGAAGACATGTTATGGAAGTTC 169
 Db 165 CAATACGTCACACTTAACTGTAATCCATTTGGAAGACATGTTATACGTGAAGGCT 224
 Qy 170 TTTATGATCAAGAGTGGCCGTAATGATGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 229
 Db 225 TATACGATCAGCAGAGATGCCAAATGATGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 284
 Qy 230 TTCCATTTGATTCATGCAATGTTGGCGGTACACGATCTCTGAATCCACGTGATTTTGG 289
 Db 285 TTCCGTTTGATTCGTAATGATGACACGCTTCGTTAATCCACGTGATTTTGG 344
 Qy 290 TAACAACAACGTGTGTGATTTGCTTTCATTCATTCATTTGTAACAAAGTTGATTCGCT 349
 Db 345 TCACATCAGTTGTGTGATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 404
 Qy 350 ATCGATGCAATGCTTTTACATGATGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 409
 Db 405 ATCGATGCAATGCTTTTACATGATGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 464
 Qy 410 -CTGAATACACATGCTTTTCAACTCAATATGT-CCCGATGCCAGTA-TGCCGTTATGA 467
 Db 465 CCCGAAATGACACATGCTTTTCAACTCAATATGT-CCCGATGCCAGTA-TGCCGTTATGA 524
 Qy 468 AATTTGGATGGTGGACCAACCGGTC 493
 Db 525 GATTCGTGATGGTGGACCATCCCGGC 550

RESULT 26
 A1317885 615 bp mRNA linear EST 17-DEC-1998
 LOCUS SMOV3MCA06H12SK Onchocerca volvulus molting L3 larva cDNA
 DEFINITION (SL96MLM-Ovml3) Onchocerca volvulus cDNA clone SMOV3MCA06H12 5',
 mRNA sequence.
 ACCESSION A1317885
 VERSION A1317885.1 GI:4033152
 KEYWORDS EST.
 SOURCE Onchocerca volvulus.
 ORGANISM Onchocerca volvulus.
 Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
 Onchocercidae; Onchocerca.

REFERENCE 1 (bases 1 to 615)
 AUTHORS Williams, S.A., Lizotte-Waniewski, M., Laney, S. and Lustigman, S.
 TITLE Genes expressed in molting L3 larvae of Onchocerca volvulus
 JOURNAL Unpublished (1997)
 COMMENT Contact: Steven A. Williams
 Molecular Parasitology
 Smith College Department of Biological Sciences
 Department of Biological Sciences, Clark Science Center, Smith
 College, Northampton, MA, 01063, USA


```

/!ab host="DH10B"
/!note="vector: pAMP1 (Gibco); Site_1: NotI; Site_2: SalI;
The library was constructed by Brandt Chapell and Dr.
James McCarter at Washington University, St. Louis. The
cDNA was made by using Dynabead oligo-dT priming (Dyna).
The PCR based library using a modified protocol from the
SMART PCR cDNA Synthesis Kit from Clontech. Nematodes were
cloned into the DPG sites of pAMP1. Nematodes were
provided by Dr. David Bird and Daniel Snyder of North
Carolina State University."
BASE COUNT      158 a      122 c      111 g      143 t
ORIGIN

```

Query Match	25.2%;	Score 292.6;	DB 13;	Length 534;
Best Local Similarity	-76.0%;	Pred. No. 9.8e-71;		
Matches 361; Conservative	0;	Mismatches 114;	Indels 0;	Gaps 0;

QY	403	GAGGATATCGAAATACACAGATCGTTTTCAAATCGAAATGTCGGCATGCAATATGCCGT	462
Db	1	GGGGGTCCGGAATATACACACAAATTCACAAACAAATTTGCTATGCTGTATGTAAA	60
QY	463	TATGAATTTTGGATGATGGAGAACCAACGGTCAACCAAGTTCAATTTGCTATCATGGTACG	522
Db	61	TATGATGTCCTCGAAGGGGGGGGGGGCTGGTCAACCAATTCATTTGGCATTTATGGCCA	120
QY	523	CCAGTTTTCATTAATGACATGCGATTTCTGAACCGTTGATCTTTCTGGCCGGTTGTC	582
Db	121	CAAGTTTATCAAAATGACCTGTGATTTCAGAGACAGTAGACACATTCGCGCTGTGTGA	180
QY	583	CATTTCCTCTTGTGTGATGATAGGTAAACGGTGATCTACTGTGGAAATTTCTAATGCTGATGA	642
Db	181	CATTCTTCTCTGTTGATGATGATGTATGTTGATCTACTGTCCAAATATTAAACGAACAGGGT	240
QY	643	TGTGCTCTTGATTAATTAATTTGCTAAATAATTTGGAATATCCACAGATTTAATGCTGGC	702
Db	241	TGTGCTTTGGATTAATTTCTTATTTAAACAATTGGAAATATCCAACTGATTTAATGCTGTGA	300
QY	703	CAAGAAGCTCACAGTATACAAATATGCGGATGCATACAGCTTTCTATCAATGCCAGATC	762
Db	301	CAAGAAGCTCAAGTTTATTAATATATCTGATCTGTTCACAACTTTTCTATCAATGCCAAATT	360
QY	763	AGTATTTACCAATTAAAGAACCAATATGCGAGATGTGTTGACACCAATGTTCAGAACCCACA	822
Db	361	AGTATTTACCAATTAAAGAACCTCACAGTGAATGTGCACGTCCAAAATGTGCAAGACTTCT	420
QY	823	GGATTCGAGCGGTAAAGAACAGGTGGTGGCCGACAAAAACCTGTCGACACTGCGC	877
Db	421	GGCCTTAAATGCTGCTCAAAAGTTGCTGCTGGAGGGGGGGGGGACGAGAGACACACC	475

RESULT	30
AA901444	
LOCUS	
DEFINITION	531 bp mRNA linear EST 26-MAR-1998
	SNOV3MCAM03B05 Onchocerca volvulus molting L3 larva cDNA
	(SI:196WLM-OymM3) Onchocerca volvulus cDNA clone onchit7 5' similar to
	TR:Q93115 Q93115 CUT-1-LIKE CUTICLIN PROTEIN PRECURSOR. ; , mRNA
	sequence.
ACCESSION	AA901444
VERSION	AA901444.1 GI:3037198
KEYWORDS	EST.
SOURCE	Onchocerca volvulus.
ORGANISM	Onchocerca volvulus
	Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
	Onchocercidae; Onchocerca.

REFERENCE
1 (pages 1 to 531)
Williams, S.A., Lizotte-Waniewski, M., Ianey, S., Lustigman, S., Hillier, L., Allen, M., Bowles, L., Gelsel, S., Jost, S., Kucha, T., Martin, J., Stepien, M., Theising, B., White, Y., Wylie, R., Chappell, J., Person, B., Gibbons, M., Harvey, N., Pape, D., Chamberlain, A., Morales, R., Schnur, R., Riller, E., Kohn, S., Underwood, K. and Marra, M.
Molecular Parasitology Ovml3
Unpublished (1998)

COMMENT

Contact: Steven A. Williams
Molecular Parasitology
Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith
College, Northampton, MA, 01063, USA
Tel: 4135853826

FEATURES
 Email: genomesmith.edu
 The library was constructed by Sara Lustigman and Michelle Lizotte-Waniewski in the Laboratory of Dr. S.A.Williams. The library is available from Dr. Sara Lustigman email slustigman@pub.crc.org When requesting this clone from Dr. Lustigman please reference the Williams lab clone id - SMO33MCAM03B05
 Seq primer: -40m13 fwd. ET from Amersham
 High quality sequence stop: 484.
 Location/Qualifiers

```

FEATURES
SOURCE
Location/Qualifiers
1. 531
/organism="Onchocerca volvulus"
/strain="Kumba, Cameroons"
/db_xref="taxon:6282"
/clone="onch17"
/clone_1ib="Onchocerca volvulus molting L3 larva cDNA
(S196MW-OvML3)"
/dev_stage="molting L3"
/lab_host="XLI-Blue MRF"
/note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:
Xho I; Filarial nematode parasite of humans. Third-stage
larvae, L3, were isolated from infected black flies in
Cameroon (forest strain). The L3 were cultured in 20% FCS
in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in
culture. L3 of O. volvulus molt to fourth-stage larvae by
day 5 in culture. mRNA was isolated from approximately
6000 molting larvae (ml3), 2000 larvae from day 1, 2 or 3
in culture, and converted to double-stranded cDNA using
reverse transcriptase and oligo(dt) followed by RNase H
and DNA pol I. The library was constructed in the lambda
Uni-Zap XR vector and has 1 x 10E6 independent
recombinants and the average insert size is ~1200 bp. The
library was constructed by Sara Lustigman and Michelle
Lizotte-Waniewski in the Laboratory of Dr. S. A. Williams.
The library is available from Dr. Sara Lustigman (email:
slustig@nbc.org)."
BASE COUNT      163 a      100 c      100 g      166 t      2 others
ORIGIN
Query Match      25.1%; Score 291.2; DB 9; Length 531;
Best Local Similarity 83.5%; Pred. No. 2.4e-70;
Matches 329; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

```

QY	50	ATTGCATTCCGGTTGACACATGTTGCGAAGGTGACCGAAGAAATTGAATGGACCAACTT	109
Db	137	ATGCTATTCCGGTTGATACGGTGTAGAAGGTGAACCGAATAATTGAATGGTCCCAACT	196
QY	110	CAATPAACAAATATTTTAATACAGCTAATGCATTCGAAAGACATGTTTATGGAAGATC	169
Db	197	CAATTACGTCAACTTTTATACTGCTATTCATTCCATTTGGAAGACATGTATACGGAAGGCT	256
QY	170	TTTATGATCAAGAAAGTTGCCGTATGATGAAGGTGACAGTTCMAAGTTGCCGAATTTTAC	229
Db	257	TATACGATCAGGACGAGATGCGNNATGATGAAGGTGGAAGCTCAGGTAGCGGGAATTTGAAC	316
QY	230	TTTCATTGGANTCATGCAATGTTGGCGGTACACGATCTCTGATCCACGCTGTAATTTTGG	289
Db	317	TTCCGTTGATTCCTGTGAATGTAGACACGTACACGTTTCTTAATCCAGCTGGTATTTTGG	376
QY	290	TAAACAACACGTGTGCTATTTCGTTTCAATTAATTTGTTACCAAGTTGATCTGCAT	349
Db	377	TCAACATCAGTTGTGTCAATCCATCCATCTGTTTGTGCAAAAGTGTATCCAGCAT	436
QY	350	ATCGATGCATAGCTTTTACATGGAAGCTGATTAACAGGTAATGTCACACAGATGAGGAT	409
Db	437	ATCGAATACAAATGCTTTTACATGGAAGCTGATTAACACAGTTAGCCCTCAACTTTGAAGTTT	496

Db	361	CCA	363	11
RESULT	34			
LOCUS	AW313040			
DEFINITION	AW313040	558 bp	mRNA	linear
ACCESSION	AW313040			EST 24-JAN-2000
VERSION	AW313040.1	GI:6742215		
KEYWORDS	EST.			
SOURCE	Onchocerca volvulus.			
ORGANISM	Onchocerca volvulus.			
REFERENCE	Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea; Onchocercidae; Onchocerca.			
AUTHORS	1 (bases 1 to 558)			
TITLE	Williams, S.A., Lizotte-Waniewski, M., Laney, S. and Lustigman, S.			
JOURNAL	Genes expressed in molting L3 larvae of Onchocerca volvulus			
COMMENT	Unpublished (1997)			
	Contact: Steven A. Williams			

Email: genome@smith.edu
Seq primer: pbluescript SK.
location/Qualifiers
1..558
FEATURES
source

BASE COUNT
 167 a 111 c 114 g 163 t 3 others

/organism="Onchocerca volvulus"
 /strain="Kumba, Cameroons"
 /db_xref="taxon:6282"
 /clone="SMOV3MCAM38A02"
 /clone_1ib="Onchocerca volvulus molting L3 larva cDNA
 (SL96MLM-OvML3)"
 /dev_stage="molting L3"
 /lab_host="XLI-Blue MR"
 /note="Vector: Lambda Uni-Zap XR; Site_1: Eco RI; Site_2:
 Xho I; Filarial nematode parasite of humans. Third-stage
 larvae, L3, were isolated from infected black flies in
 Cameroon (forest strain). The L3 were cultured in 20% FCS
 in IMDM+NCTC 135 and collected after day 1, 2, or 3 in
 culture. L3 of O. volvulus molt to fourth-stage larvae by
 day 5 in culture. mRNA was isolated from approximately
 6000 molting larvae (ml3), 2000 larvae from day 1, 2 or 3
 in culture, and converted to double-stranded cDNA using
 reverse transcriptase and oligo(dT) followed by RNase H
 and DNA pol I. The library was constructed in the lambda
 Uni-Zap XR vector and has 1 x 10⁶ independent
 recombinants and the average insert size is ~1200 bp. The
 library was constructed by Sara Lustigman and Michelle
 Lizotte-Waniewski in the laboratory of Dr. S. A. Williams.
 The library is available from Dr. Sara Lustigman (email:
 slustigman@bc.org)."

Query Match	23.9%;	Score 277.8;	DB 10;	Length 558;
Best Local Similarity	71.9%;	Pred. No. 1.4e-66;		
Matches 388;	Conservative	0;	Mismatches 150;	Indels 2;
			Gaps	2;

QY	7	ATTCGCTTAATGCTTTTCGTGACACATTAATGCAATGT-CTTAATGAAATCCGGTGA	65
Db	17	ATCTCTCTCTCGTATTGCTACTCTTTCTACTTGACACGTGTTTCAATCCATTCCTATGA	76
QY	66	CAATGGTGTGCAAGGTGAGCAAAATGATGTGGACCACTCAATATCAATCAATTT	125
Db	77	TAAACGTGTGCAAGGTGAACTGGAATAGATGTGGCCCAACTTCGCTAAACATTCATATTT	136
QY	126	TAAATACAGCTAATGCAATTCGAAGGACATGTTTATGTGAAAGTCTTTATGATCAAGAGG	185

Db 137 CAATCTACGAAATACATTTGAAAGTCAATGATATACGTAATAAAGGCTTTACATGCGCATGA 196

QY 186 TTGCGGTAATGATGAAGGTGGACGCTCAAGTTGCCGAATTCACCTTCATTGTTTCATG 245

Db 197 ATGTCCGTTCAGATAGTGGTGGACGCGCAGGTACGTGGAATTTAAAGTTAGATGCATTCATG 256

QY 246 CAATGTTGGCGGTACACGATCTCTGTAATCCACAGTGGTAATTTTGTAAACAACATGTTG 305

Db 257 TAAATGTTGAACGATCAACGGTCCCTTAATTCCTGAGAGTGTCTTTGTATCAACACATTAAT 316

QY 306 CATTGCTTTTCATCCATTAATTTGTTTACCAGAAAGTTGATCGTCATATGAGTACATGCTT 365

Db 317 TATTTCATTTCATCCGAATTCATTACAAAATATGACGACACATATGATATACAGTGT 376

QY 366 TTTCATGGAAGGTGATATAAACAGTTAGTGCACAGATTGAGTATCTGTAATTCACAACTGC 425

Db 377 CTATATGGAAGGTGCACAAAACCGTTAGTGCAGAACTGGAAGTATCCGAATATGACCACAG 436

QY 426 TTTTCAACTCCAAATTTGTCGCCGATGCCAGTATGCCGTTATGAATTTTGGATGTGGACC 485

Db 437 ATTCCGACACAAAGTGTATCCATATCCCTGTGTGTCGATCCCAAGAAATCTGGAAAGNGAGAC 496

QY 486 AACCGGTCAACAGTTTCATTTGCTATCATTTGGTCAGC - CAGTTTATCAATAATGGACAT 544

Db 497 AACTGTTACACCTATTCGATATGCATATGTGGAGATACACGTAATATCACAAAATGGGCTT 556

RESULT 35
AW312995
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AW312995 543 bp mRNA linear EST 24-JAN-2000
SMOW3CMAM37E11SK Onchocerca volutus molting L3 larva cDNA
(SL56XMTM-Ovmu13) Onchocerca volutus cDNA clone SMOW3CMAM37E11 5',
mRNA sequence.
AW312995
AW312995.1 GI:6742181
EST.
Onchocerca volutus.
Onchocerca volutus
Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidae;
Onchocercidae; Onchocerca.
1 (bases 1 to 543)
Williams, S.A., Lizotte-Waniewski, M., Laney, S. and Lustigman, S.
Genes expressed in molting L3 larvae of Onchocerca volutus
Unpublished (1997)
Contact: Steven A. Williams
Molecular Parasitology
Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith
College, Northampton, MA, 01063, USA
Tel: 4135853826
Fax: 4135853786
Email: genome@smith.edu
Seq primer: Bluescript SK.
Location/Qualifiers
1..543
Source

```

/organism="Onchocerca volvulus"
/strain="Kumba, Cameroon"
/db_xref="taxon:6282"
/clone="SMOV3KCM37E1"
/clone_1lb="Onchocerca volvulus molting L3 larva cDNA
(SJ66MCM-Ovml3)"
/dev_stage="molting L3"
/lab_host="XLI-Blue MRF"
/note="Vector: lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:
Xho I; Filarial nematode parasite of humans. Third-stage
larvae, L3, were isolated from infected black flies in
Cameroon (forest strain). The L3 were cultured in 20% FCS
in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in
culture. L3 of O. volvulus molt to fourth-stage larvae by
day 5 in culture. mRNA was isolated from approximately
6000 molting larvae (ml3). 2000 larvae from day 1, 2 or 3
in culture, and converted to double-stranded cDNA using

```


BASE COUNT
ORIGIN

FEATURES

```

/organism="Onchocerca volvulus"
/strain="Kumba, Cameroons"
/db_xref="taxon:6282"
/clone="SMuJ3C02.020"
/clone_1lb="Onchocerca volvulus molting L3 larva cDNA
(SL96MLW-Ovml3)"
/dev_stage="molting L3"
/lab_host="Xp1-Blue MRP"
/notes="Vector: lambda uni-zap XR; Site_1: Eco RI; Site_2:
Xho I; Filarial nematode parasite of humans. Third-stage
larvae, L3, were isolated from infected black flies in
Cameroon (forest strain). The L3 were cultured in 20% FCS

```

in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in culture. L3 of *O. volvulus* molt to fourth-stage larvae by day 5 in culture. mRNA was isolated from approximately 6000 molting larvae (ML3), 2000 larvae from day 1, 2 or 3 in culture, and converted to double-stranded cDNA using reverse transcriptase and oligo(dT) followed by RNase H and DNA pol I. The library was constructed in the lambda Uni-Zap XR vector and has 1 x 10⁶ independent recombinants and the average insert size is ~1200 bp. The library was constructed by Sara Lustigman and Michelle Lizotte-Waniewski in the Laboratory of Dr. S. A. Williams. The library is available from Dr. Sara Lustigman (email: slustig@nybc.org)."

BASE COUNT 131 a 85 c 91 g 150 t 2 others
ORIGIN

Query Match 22.7%; Score 263; DB 9; Length 459;

Best Local Similarity 83.9%; Pred. No. 1.7e-62;

Matches 296; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

```
QY 50 ATTCGATTCGGGTGACATGCTGTCGAGGTGACCCAGAAATTGAATGTGACCACTT 109
    |||||||
Db 81 ATGCTATTCGGTGTGATTAAGGTGATGACGTAACCAAAATTGAATGTGTCACACT 140
    |||||||

QY 110 CAATTAACATCAATTTTAATACAGTAATGATGACGTAAGGACATGTTATGTGAAGTGC 169
    |||||
Db 141 CAATTAAGTCAACTTTAATACGTAATGATGACGTAAGGACATGTTATGTGAAGGCT 200
    |||||

QY 170 TTATGATCAAGAGGTGGCGTAATGATGAGGTGACGTCATCAAGTCCGGAATTTTAC 229
    |||||||
Db 201 TATACGATCAGCAGGATGCCGAATGATGAGGTGACGTCAGTACCGGAATTTGAC 260
    |||||||

QY 230 TTCCATTTGATTCATGATGATGCTGCGGTACAGATCTGAAATCCAGTGTATTTTGG 289
    |||||||
Db 261 TTCCGTTTGTCTCTGAATGTANCAAGTTCGTTAAATCCAGTGTATTTTGG 320
    |||||||

QY 290 TAACAACAAGTGTGTCATTTGTCATTCATTTATTTTACCAAGTGTGATGTCAT 349
    |||||
Db 321 TCACATCAGTGTGTCATTCATTCATTCATTTGTCACAAAAGTGTGATGACGAT 380
    |||||

QY 350 ATCGAGTACAGTCTTTTACATGGAAGCTGATAAACAGTTAGTGACAGATT 402
    |||||
Db 381 ATCGAATACAGTCTTTTACATGGAAGCTGATAAACAGTTAGTGACGCTCAACT 433
    |||||
```

Search completed: March 28, 2003, 16:58:33
Job time : 1655 secs

GenCore version 5.1.4 p5 4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 24, 2003, 08:49:12 ; Search time 41 Seconds
(without alignments)
1257.736 Million cell updates/sec

Title: US-10-054-562a-4

Perfect score: 2026

Sequence: 1 MMRLIAFCTLTIALSYSIP.....ALIAAVITTSFRRPNQKA 387

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq_101002.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2026	100.0	387	22	AA84258	Amino acid sequence
2	2026	100.0	387	23	AB876278	Dirofilaria immitis
3	1086.5	53.6	245	22	AA84260	Amino acid sequence
4	1086.5	53.6	245	23	AB876280	Brugia malayi cuti
5	776.5	38.3	271	22	AA84259	Amino acid sequence
6	776.5	38.3	271	23	AB876279	Dirofilaria immitis
7	199	9.8	462	22	AB859975	Drosophila melanog
8	166	8.2	604	22	AB870061	Drosophila melanog
9	163	8.0	1638	22	AB867889	Drosophila melanog
10	162.5	8.0	611	22	AB867239	Drosophila melanog

11	162.5	8.0	611	22	AB870054	Drosophila melanog
12	150	7.4	758	22	AB858337	Drosophila melanog
13	142	7.0	833	22	AB858185	Drosophila melanog
14	140	6.9	692	22	AB870355	Drosophila melanog
15	135.5	6.7	744	22	AB858426	Drosophila melanog
16	130.5	6.4	638	22	AB864380	Drosophila melanog
17	124.5	6.1	699	22	AB864367	Drosophila melanog
18	119.5	5.9	418	22	AB868164	Drosophila melanog
19	111	5.5	2284	22	AB871444	Drosophila melanog
20	100.5	5.0	768	22	AB868660	Drosophila melanog
21	100	4.9	935	13	AB820112	AE-III (peptidylhy
22	97	4.8	227	22	AB871814	Drosophila melanog
23	94.5	4.7	693	10	AB894856	Expression plasmid
24	94.5	4.7	875	10	AB894854	C-terminal prepro-
25	94.5	4.7	875	16	AB873053	Peptidyl C-termina
26	93	4.6	474	22	AB863130	Drosophila melanog
27	92	4.5	896	23	AB808760	Synechococcus cyan
28	90	4.4	2515	22	AB864427	Drosophila melanog
29	87.5	4.3	474	21	AA870900	Protein encoded by
30	87.5	4.3	474	21	AA870923	Human soluble CD39
31	87	4.3	473	21	AA870901	Protein encoded by
32	87	4.3	473	21	AA870924	Human soluble CD39
33	87	4.3	3038	16	AA874171	Aspergillus terreu
34	86.5	4.3	636	22	AA894352	Human protein sequ
35	86.5	4.3	747	22	AA840070	Human polypeptide
36	86.5	4.3	747	22	AA893124	Human protein sequ
37	85.5	4.2	438	23	AB838284	Staphylococcus epl
38	85.5	4.2	560	18	AA813009	Segment of desmoso
39	85.5	4.2	940	22	AA839436	Human polypeptide
40	85.5	4.2	1117	23	AA878053	Human desmoglein 2
41	85.5	4.2	1118	23	AA820266	Human lung specifl
42	85.5	4.2	1120	22	AB810284	Novel human diagno
43	85.5	4.2	1121	22	AA841282	Human polypeptide
44	85	4.2	2069	22	AA830566	C glutamicum prote
45	84	4.1	474	22	AA894481	Human protein sequ

ALIGNMENTS

RESULT 1
AA84258
ID AAB84258 standard; Protein: 387 AA.
XX
AC AAB84258;
XX
DT 22-AUG-2001 (first entry)
XX
DE Amino acid sequence of a cuticlin polypeptide.
XX
KW Cuticlin; gene therapy; vaccine; helminth parasite.
XX
OS Dirofilaria immitis.
XX
PN US6248329-B1.
XX
PD 19-JUN-2001.
XX
PF 01-JUN-1999; 99US-0323427.
XX
PR 01-JUN-1998; 98US-0087435.
XX
PA (CHAN/) CHANDRASHEKAR R.
PA (MORA/) MORALES T H.
PI Chandrashekar R, Morales TH;
XX
XX WPI: 2001-396953/42.
DR N-PSDB: AAH24826, AAH24828.
XX
PT Novel nucleic acid molecule encoding a Dirofilaria immitis cuticlin
PT protein is useful as a vaccine to prevent parasitic helminth infection

XX Claim 3: Column 39-40; 29pp; English.
PS
XX
CC The specification describes a *Dirofilaria immitis* nucleic acid molecule,
CC that encodes a cuticlin protein. The *Dirofilaria immitis* nucleic acid
CC molecule is useful as a probe to identify nucleic acid molecules, as a
CC primer to produce nucleic acids, as a therapeutic reagent, e.g., gene
CC therapy to inhibit cuticlin activity or production, or in a vaccine to
CC prevent infection with helminth parasites. The cuticlin protein,
CC antibodies raised against it, and inhibitory compounds of cuticlin may
CC all be used in compositions to protect animals, especially mammals such
CC as cats, dogs, and humans. The antibodies may be used to passively
CC immunize an animal, or as reagents in assay to detect infection of
CC helminths, or as tools to screen expression libraries to recover desired
CC proteins. They may also be used to target cytotoxic agents to the
CC parasite and kill it directly. The present sequence represents a cuticlin
CC polypeptide.
XX
SO Sequence 387 AA:
Query Match 100.0%; Score 2026; DB 22; Length 387;
Best Local Similarity 100.0%; Pred. No. 7.1e-210;
Matches 387; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MMRLIAFCTTLALSYSPVNGVGEPEIECGPTSTINFNTRNAFGHYVVGGLYDQ 60
DB 1 MMRLIAFCTTLALSYSPVNGVGEPEIECGPTSTINFNTRNAFGHYVVGGLYDQ 60
QY 61 ECGRNDGGRQVAGISLPDSCNVARTSLNPRGIFVTTTVISFHPLEFVKVDRARYQ 120
DB 61 ECGRNDGGRQVAGISLPDSCNVARTSLNPRGIFVTTTVISFHPLEFVKVDRARYQ 120
QY 121 CFYMEADKTVSAQIEVSETTAFQTQIYMPVCRYEILDGPGTQPVOPAIIGQPYHKM 180
DB 121 CFYMEADKTVSAQIEVSETTAFQTQIYMPVCRYEILDGPGTQPVOPAIIGQPYHKM 180
QY 181 TCDSEVDTFCAVHSCFVDDNGDVTVEILNADGALDKYLLNNLEYPTDLMAAGEAHY 240
DB 181 TCDSEVDTFCAVHSCFVDDNGDVTVEILNADGALDKYLLNNLEYPTDLMAAGEAHY 240
QY 241 KYADRSQLFYQCOISITTIKEPNSCEVRPQCSPOGFGAVKTGGAAPAAAQRLRLKKR 300
DB 241 KYADRSQLFYQCOISITTIKEPNSCEVRPQCSPOGFGAVKTGGAAPAAAQRLRLKKR 300
QY 301 SAEPENIIDVRDINTLEISDNOALPVDLRHALLQHNQGVYIIAAVONGICMSPFGFS 360
DB 301 SAEPENIIDVRDINTLEISDNOALPVDLRHALLQHNQGVYIIAAVONGICMSPFGFS 360
QY 361 MEMGLSIALIAVAVITISFKFRPNOKA 387
DB 361 MEMGLSIALIAVAVITISFKFRPNOKA 387
RESULT 2
ID ABB76278 standard; Protein; 387 AA.
XX ABB76278;
AC
XX
DT 12-AUG-2002 (first entry)
XX
DE *Dirofilaria immitis* cuticlin DiCut-1A.
XX
XX Cuticlin; helminth; parasite; DiCut-1A; vaccine; gene therapy;
KW anthelmintic.
XX
OS *Dirofilaria immitis*.
XX
PN US2002037294-A1.
XX
PD 28-MAR-2002.
XX
PF 20-MAR-2001; 2001US-0812642.

XX
PR 01-JUN-1999; 99US-0323427.
XX
XX (CHAN/) CHANDRASHEKAR R.
PA (MORA/) MORALES T H.
XX
PI Chandrashekar R, Morales TH;
DR WPI; 2002-401306/43.
DR N-PSDB; ABL57372.
XX
PS Claim 4: Page 21-22; 32pp; English.
XX
CC The present sequence is the protein sequence of novel
CC *Dirofilaria immitis* cuticlin DiCut-1A, as predicted from an
CC isolated cDNA clone (see ABL57372). The protein sequence has 91%
CC identity to the *Ascaris* homologue. The invention provides D
CC *immitis* and *Brugia malayi* parasitic helminth cuticlin proteins,
CC nucleic acids, antibodies, compounds that inhibit cuticlin
CC activity, and methods of obtaining them. Therapeutic compositions
CC comprising such proteins, nucleic acids, antibodies and/or
CC inhibitors, including genetic vaccines, recombinant virus vaccines
CC and recombinant cell vaccines, are used to protect animals from
CC diseases caused by parasitic helminths, and to inhibit the moulting
CC of filarid larvae in an animal.
XX
SO Sequence 387 AA:
Query Match 100.0%; Score 2026; DB 23; Length 387;
Best Local Similarity 100.0%; Pred. No. 7.1e-210;
Matches 387; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MMRLIAFCTTLALSYSPVNGVGEPEIECGPTSTINFNTRNAFGHYVVGGLYDQ 60
DB 1 MMRLIAFCTTLALSYSPVNGVGEPEIECGPTSTINFNTRNAFGHYVVGGLYDQ 60
QY 61 ECGRNDGGRQVAGISLPDSCNVARTSLNPRGIFVTTTVISFHPLEFVKVDRARYQ 120
DB 61 ECGRNDGGRQVAGISLPDSCNVARTSLNPRGIFVTTTVISFHPLEFVKVDRARYQ 120
QY 121 CFYMEADKTVSAQIEVSETTAFQTQIYMPVCRYEILDGPGTQPVOPAIIGQPYHKM 180
DB 121 CFYMEADKTVSAQIEVSETTAFQTQIYMPVCRYEILDGPGTQPVOPAIIGQPYHKM 180
QY 181 TCDSEVDTFCAVHSCFVDDNGDVTVEILNADGALDKYLLNNLEYPTDLMAAGEAHY 240
DB 181 TCDSEVDTFCAVHSCFVDDNGDVTVEILNADGALDKYLLNNLEYPTDLMAAGEAHY 240
QY 241 KYADRSQLFYQCOISITTIKEPNSCEVRPQCSPOGFGAVKTGGAAPAAAQRLRLKKR 300
DB 241 KYADRSQLFYQCOISITTIKEPNSCEVRPQCSPOGFGAVKTGGAAPAAAQRLRLKKR 300
QY 301 SAEPENIIDVRDINTLEISDNOALPVDLRHALLQHNQGVYIIAAVONGICMSPFGFS 360
DB 301 SAEPENIIDVRDINTLEISDNOALPVDLRHALLQHNQGVYIIAAVONGICMSPFGFS 360
QY 361 MEMGLSIALIAVAVITISFKFRPNOKA 387
DB 361 MEMGLSIALIAVAVITISFKFRPNOKA 387
RESULT 3
ID AAB84260 standard; Protein; 245 AA.
XX AAB84260;
AC
XX
DT 22-AUG-2001 (first entry)


```

XX      Amino acid sequence of a cuticlin polypeptide.
DE
XX      Cuticlin: gene therapy; vaccine; helminth parasite.
KM
XX      Brugia malayi.
OS
XX      US6248329-B1.
PN
XX      19-JUN-2001.
PD
XX      01-JUN-1999; 99US-0323427.
PE
XX      01-JUN-1998; 98US-0087435.
PR
XX      (CHAN/) CHANDRASHEKAR R.
PA      (MORA/) MORALES T H.
PI      Chandrashekar R, Morales TH;
XX      WPI: 2001-396953/42.
DR      N-PSDB; AAB24839.
XX
XX      Novel nucleic acid molecule encoding a Dirofilaria immitis cuticlin
PT      protein is useful as a vaccine to prevent parasitic helminth infection
PT
XX      Disclosure; Column 53-54; 29pp; English.
PS
XX
XX      The specification describes a Dirofilaria immitis nucleic acid molecule,
CC      that encodes a cuticlin protein. The Dirofilaria immitis nucleic acid
CC      molecule is useful as a probe to identify nucleic acid molecules, as a
CC      primer to produce nucleic acids, as a therapeutic reagent, e.g., gene
CC      therapy to inhibit cuticlin activity or production, or in a vaccine to
CC      prevent infection with helminth parasites. The cuticlin protein,
CC      antibodies raised against it, and inhibitory compounds of cuticlin may
CC      all be used in compositions to protect animals, especially mammals such
CC      as cats, dogs, and humans. The antibodies may be used to passively
CC      immunize an animal, or as reagents in assay to detect infection of
CC      helminths, or as tools to screen expression libraries to recover desired
CC      proteins. They may also be used to target cytotoxic agents to the
CC      parasite and kill it directly. The present sequence represents a cuticlin
CC      polypeptide.
XX
SQ      Sequence 245 AA;
Query Match 53.6%; Score 1086.5; DB 22; Length 245;
Best Local Similarity 80.6%; Pred. No. 1.3e-108;
Matches 195; Conservative 24; Mismatches 22; Indels 1; Gaps 1;
QY      3 IRLAFCITLIALSY-SIPVDNGVEGEPEICGPTSTINNTNNAFEGHYVVGLYDOE 61
DB      4 MOICSFSLYMIASINMIPIDNGVESPEICGPTSTIVNTNPNPEGHYAGLYSNQ 63
QY      62 GCRNDEGROVAGISLPEDSCNVAARTSLNRGIFVTTVVISFHPLEFVTVDRAYVQC 121
DB      64 DCRSDEGROVAGISLPEDSCNVAARTSLNRGIFVTAVVITTHPQGITVDRYRQC 123
QY      122 FYMEADKTVAQIEVSEITTAFOQIYVMPVCRREILDGPTGQPVQYANIGOPVYHKWT 181
DB      124 FYMEADKTVAQIEVSEITTVFAQIYVMPVCRREILDGPTGQPVQYANIGOPVYHKWT 183
QY      182 CDSETVDFCAVHSCFVDGNGDTVELNADGALDKYLLNNLEYPTDLMAAGBAHYK 241
DB      184 CDSETVDFCALVHSCFVDGNGDSINLINDEGALDRYLLNNLEYPTDLMAAGBAHYK 243
QY      242 YA 243
DB      244 YA 245

```

```

ID      ABB76280 standard; Protein; 245 AA.
XX
XX      ABB76280;
AC
XX      12-AUG-2002 (first entry)
DT
XX      Brugia malayi cuticlin BmCut-1A.
DE
XX      Cuticlin; helminth; parasite; BmCut-1A; vaccine; gene therapy;
KM      anthelmintic.
XX
XX      Brugia malayi.
OS
XX      US2002037294-A1.
PN
XX      28-MAR-2002.
PD
XX      20-MAR-2001; 2001US-0812642.
PE
XX      01-JUN-1999; 99US-0323427.
PR
XX      (CHAN/) CHANDRASHEKAR R.
PA      (MORA/) MORALES T H.
PI      Chandrashekar R, Morales TH;
XX      WPI: 2002-401306/43.
DR      N-PSDB; ABL57374.
XX
XX      New parasitic helminth cuticlin proteins and nucleic acids isolated
PT      from Dirofilaria immitis, useful for protecting animals from diseases
PT      caused by parasitic helminths or for inhibiting molting of filarid
PT      larvae in an animal.
XX
XX      Example 1; Page 28-29; 32pp; English.
PS
XX
XX      The present sequence is the protein sequence of novel
CC      Brugia malayi cuticlin BmCut-1A, as predicted from an isolated
CC      partial cDNA clone (see ABL57374). The invention provides B.
CC      malayi and Dirofilaria immitis parasitic helminth cuticlin proteins,
CC      nucleic acids, antibodies, compounds that inhibit cuticlin
CC      activity, and methods of obtaining them. Therapeutic compositions
CC      comprising such proteins, nucleic acids, antibodies and/or
CC      inhibitors, including genetic vaccines, recombinant virus vaccines
CC      and recombinant cell vaccines, are used to protect animals from
CC      diseases caused by parasitic helminths, and to inhibit the molting
CC      of filarid larvae in an animal.
XX
SQ      Sequence 245 AA;
Query Match 53.6%; Score 1086.5; DB 23; Length 245;
Best Local Similarity 80.6%; Pred. No. 1.3e-108;
Matches 195; Conservative 24; Mismatches 22; Indels 1; Gaps 1;
QY      3 IRLAFCITLIALSY-SIPVDNGVEGEPEICGPTSTINNTNNAFEGHYVVGLYDOE 61
DB      4 MOICSFSLYMIASINMIPIDNGVESPEICGPTSTIVNTNPNPEGHYAGLYSNQ 63
QY      62 GCRNDEGROVAGISLPEDSCNVAARTSLNRGIFVTTVVISFHPLEFVTVDRAYVQC 121
DB      64 DCRSDEGROVAGISLPEDSCNVAARTSLNRGIFVTAVVITTHPQGITVDRYRQC 123
QY      122 FYMEADKTVAQIEVSEITTAFOQIYVMPVCRREILDGPTGQPVQYANIGOPVYHKWT 181
DB      124 FYMEADKTVAQIEVSEITTVFAQIYVMPVCRREILDGPTGQPVQYANIGOPVYHKWT 183
QY      182 CDSETVDFCAVHSCFVDGNGDTVELNADGALDKYLLNNLEYPTDLMAAGBAHYK 241
DB      184 CDSETVDFCALVHSCFVDGNGDSINLINDEGALDRYLLNNLEYPTDLMAAGBAHYK 243
QY      242 YA 243
DB      244 YA 245

```

```

RESULT 5
ID AAB84259 standard; Protein: 271 AA.
XX
XX AAB84259;
XX
XX
XX 22-AUG-2001 (first entry)
XX
XX Amino acid sequence of a cuticlin polypeptide.
XX
XX Cuticlin: gene therapy; vaccine; helminth parasite.
XX
XX Dirofilaria immitis.
XX
XX US6248329-B1.
XX
XX 19-JUN-2001.
XX
XX
XX 01-JUN-1999; 99US-0323427.
XX
XX 01-JUN-1998; 98US-0087435.
XX
XX (CHAN/) CHANDRASHEKAR R.
XX (MORA/) MORALES T H.
XX
XX Chandrashekar R, Morales TH;
XX
XX MPI; 2001-396953/42.
XX N-PSDB; AAH24830, AAH24832.
XX
XX Novel nucleic acid molecule encoding a Dirofilaria immitis cuticlin
XX protein is useful as a vaccine to prevent parasitic helminth infection
XX
XX
XX Example 1; Column 47-50; 29pp; English.
XX
XX The specification describes a Dirofilaria immitis nucleic acid molecule,
XX that encodes a cuticlin protein. The Dirofilaria immitis nucleic acid
XX molecule is useful as a probe to identify nucleic acid molecules, as a
XX primer to produce nucleic acids, as a therapeutic reagent, e.g., gene
XX therapy to inhibit cuticlin activity or production, or in a vaccine to
XX prevent infection with helminth parasites. The cuticlin protein,
XX antibodies raised against it, and inhibitory compounds of cuticlin may
XX all be used in compositions to protect animals, especially mammals such
XX as cats, dogs, and humans. The antibodies may be used to passively
XX immunize an animal, or as reagents in assay to detect infection of
XX helminths, or as tools to screen expression libraries to recover desired
XX proteins. They may also be used to target cytotoxic agents to the
XX parasite and kill it directly. The present sequence represents a cuticlin
XX polypeptide.
XX
XX Sequence 271 AA;
XX
XX
XX Query Match 38.3%; Score 776.5; DB 22; Length 271;
XX Best Local Similarity 72.4%; Pred. No. 4.8e-75;
XX Matches 144; Conservative 24; Mismatches 22; Indels 9; Gaps 2;
XX
XX 134 IEVSEITTAFOQIYVPMPCRYEILIDGPTGQPVQFAITGQPVYHKWTCDSSETVDFCAV 193
XX :||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
XX 15 LEVSEMTTAFQIYVPMPCRYEILIDGPTGQPVQFAITGQPVYHKWTCDSSETVDFCAV 74
XX
XX 194 VHSCEFDDNGDPTVELNADGALDKYLLNNLEYPTDLMAQGEAHVYKXADRSOLFYOQ 253
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 75 VHSCEVDDGKGAVALINEEGCALDKYLLNNLEYITDLMAQGEAHVYKXADRSOLFYOQ 134
XX
XX 254 ISITTEKPESECVPCQSEPOGFAVKTG-----GAAKPPAAAL---RLTKRSAEP 304
XX ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 135 ISITTEKPESECVPCQSEPOGFAVKTG-----GAAKPPAAAL---RLTKRSAEP 304
XX ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 305 ENIIDVTRDINTLEISDN 323
XX :||| | :|||:|

```

```

DB 195 DNTVDVSTGFSTVDITEEN 213
XX
XX RESULT 6
XX ABB76279
XX ID ABB76279 standard; Protein: 271 AA.
XX
XX ABB76279;
XX
XX 12-AUG-2002 (first entry)
XX
XX Dirofilaria immitis cuticlin DiCut-1B.
XX
XX Cuticlin: helminth; parasite; DiCut-1B; vaccine; gene therapy;
XX anthelmintic.
XX
XX Dirofilaria immitis.
XX
XX US2002037294-A1.
XX
XX 28-MAR-2002.
XX
XX 20-MAR-2001; 2001US-0812642.
XX
XX 01-JUN-1999; 99US-0323427.
XX
XX (CHAN/) CHANDRASHEKAR R.
XX (MORA/) MORALES T H.
XX
XX Chandrashekar R, Morales TH;
XX
XX MPI; 2002-401306/43.
XX N-PSDB; ABL57373.
XX
XX New parasitic helminth cuticlin proteins and nucleic acids isolated
XX form Dirofilaria immitis, useful for protecting animals from diseases
XX caused by parasitic helminths or for inhibiting molting of filarid
XX larvae in an animal.
XX
XX Claim 4; Page 25-26; 32pp; English.
XX
XX The present sequence is the protein sequence of novel
XX Dirofilaria immitis cuticlin DiCut-1B, as predicted from an
XX isolated cDNA clone (see ABL57373). The protein sequence has 81%
XX identity to the Ascaris homologue. The invention provides D.
XX immitis and Brugia malayi parasitic helminth cuticlin proteins,
XX nucleic acids, antibodies, compounds that inhibit cuticlin
XX activity, and methods of obtaining them. Therapeutic compositions
XX comprising such proteins, nucleic acids, antibodies and/or
XX inhibitors, including genetic vaccines, recombinant virus vaccines
XX and recombinant cell vaccines, are used to protect animals from
XX diseases caused by parasitic helminths, and to inhibit the molting
XX of filarid larvae in an animal.
XX
XX Sequence 271 AA;
XX
XX
XX Query Match 38.3%; Score 776.5; DB 23; Length 271;
XX Best Local Similarity 72.4%; Pred. No. 4.8e-75;
XX Matches 144; Conservative 24; Mismatches 22; Indels 9; Gaps 2;
XX
XX 134 IEVSEITTAFOQIYVPMPCRYEILIDGPTGQPVQFAITGQPVYHKWTCDSSETVDFCAV 193
XX :||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
XX 15 LEVSEMTTAFQIYVPMPCRYEILIDGPTGQPVQFAITGQPVYHKWTCDSSETVDFCAV 74
XX
XX 194 VHSCEFDDNGDPTVELNADGALDKYLLNNLEYPTDLMAQGEAHVYKXADRSOLFYOQ 253
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 75 VHSCEVDDGKGAVALINEEGCALDKYLLNNLEYITDLMAQGEAHVYKXADRSOLFYOQ 134
XX
XX 254 ISITTEKPESECVPCQSEPOGFAVKTG-----GAAKPPAAAL---RLTKRSAEP 304
XX ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 135 ISITTEKPESECVPCQSEPOGFAVKTG-----GAAKPPAAAL---RLTKRSAEP 304
XX ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 305 ENIIDVTRDINTLEISDN 323
XX :||| | :|||:|

```

Db 195 DNTVDSTGFSTVDITEEN 213

RESULT 7
ID ABB59975 standard; Protein; 462 AA.

AC ABB59975;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 6717.

KM Drosophila; developmental biology; cell signalling; insecticide;
pharmaceutical.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PE 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

PA (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

PI MPI: 2001-656860/75.

DR N-PSDB; ABL04078.

PT New isolated nucleic acid detection reagent for detecting 1000 or more
genes from Drosophila and for elucidating cell signalling and cell-cell
interactions -

PS Disclosure; SEQ ID NO 6717; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent
capable of detecting 1000 or more genes from Drosophila. The invention is
useful in developmental biology and in elucidating cell signalling and
cell-cell interactions in higher eukaryotes for the development of
insecticides, therapeutics and pharmaceutical drugs. The invention
discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
sequences (AB101840-AB16175) and the encoded proteins
(ABBS7737-ABB72072).

CC The sequence data for this patent did not form part of the printed
specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pct_sequences.

CC Sequence 462 AA;

Query Match 9.8%; Score 199; DB 22; Length 462;

Best Local Similarity 22.2%; Pred. No. 3e-12;

Matches 94; Conservative 55; Mismatches 150; Indels 124; Gaps 20;

QY 25 VEGEPETEC--GPTSTI-----NENTRNAPEGHVVYKGLYDOGCRNDEGGQVAG--- 74

DB 59 VEPSVRKICLGSMLITIKAPPNHET-GFSGMITYPKGLSKNSTCLSEY--RDHVGSLR 115

QY 75 ISLPDSCNVARIRSLNPRGIFVTYVVISFHPLEFVKVDRAYRVOCFYMEADKTV---- 130

DB 116 YKLPLNSCN-TMRETDGSGIEFPNTVLDPLHLKLITDLGRGHVRCAYSRDAAMPKK 174

QY 131 -----SAQIEVSEITTAFOQTIVPM 150

DB 175 YLRKHAKQPAFRSDRREYGRSLDKQDDDLDEEDYDANAPRQEDVT-----NNIIPM 230

QY 151 PVC-----RYELDGGPRGQPVQFAITGQPVYHKWTCSEYVDTFCAYVHSCFVDGN 203

Db 231 PGCHMKIYNDEHKIADVDKIGDPLTIYI-----SIDKQKV--YGLHVTDCIVRDGL 279

QY 204 G-DTVEILNADGALDXYLLNNLEYPTDLMAGQ---EAHVYKVDARSOLFQCOISITIK 259

Db 280 GWGQRLVGEDGCPMDNEIKQGFNTYDRLAANTFPFAHKFPYT--TSYIQCWVRICAL 337

QY 260 EPNSECVRPQCSSEPQGFAGVKTGGAAKPAAPAOIRLLKRSAPENI---IDVRTDINT 316

Db 338 EDPICQEPQCS-----GKRPKROAAD-----SKREDGLPATIEFSGIYV 379

QY 317 LEISDDNOLPVDLRHALLQHNQCPVITLAAVONGICSPFGESMFNG-----LSIALIA 371

Db 380 NENENAN-----DSDEDAVYKEK-----TLDALCVSQRFAIATAIAGLIIMAVVA 427

QY 372 AVI 374

Db 428 AVL 430

RESULT 8

ABBS70061

ID ABB70061 standard; Protein; 604 AA.

AC ABB70061;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 36975.

KM Drosophila; developmental biology; cell signalling; insecticide;
pharmaceutical.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PE 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

PA (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

PI MPI: 2001-656860/75.

DR N-PSDB; ABL14164.

PT New isolated nucleic acid detection reagent for detecting 1000 or more
genes from Drosophila and for elucidating cell signalling and cell-cell
interactions -

PS Disclosure; SEQ ID NO 36975; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent
capable of detecting 1000 or more genes from Drosophila. The invention is
useful in developmental biology and in elucidating cell signalling and
cell-cell interactions in higher eukaryotes for the development of
insecticides, therapeutics and pharmaceutical drugs. The invention
discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
sequences (AB101840-AB16175) and the encoded proteins
(ABBS7737-ABB72072).

CC The sequence data for this patent did not form part of the printed
specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pct_sequences.

CC Sequence 604 AA;

Query Match 8.2%; Score 166; DB 22; Length 604;

Best Local Similarity 22.0%; Pred. No. 1.7e-08;

Matches 72; Conservative 54; Mismatches 135; Indels 66; Gaps 17;

[illegible]

Result 9	
ABB67889	
ID	ABB67889 standard; Protein; 1638 AA.
XX	
AC	ABB67889;
XX	
DT	26-MAR-2002 (first entry)
XX	
DE	Drosophila melanogaster polypeptide SEQ ID NO 30459.
XX	
KW	Drosophila: developmental biology; cell signalling; insecticide; pharmaceutical.
KW	
OS	Drosophila melanogaster.
XX	
PN	MO200171042-A2.
XX	
PD	27-SEP-2001.
XX	
PF	23-MAR-2001; 2001WO-US09231.
XX	
PR	23-MAR-2000; 2000US-191637P.
PR	11-JUL-2000; 2000US-0614150.
XX	
PA	(PEKE) PE CORP NY.
XX	
PI	Venter JC, Adams M, Li PMD, Myers EM;
XX	
DR	WPI: 2001-656860/75.
DR	N-PSDB; ABL11992.
PT	
PT	New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -
XX	
PS	Disclosure; SEQ ID NO 30459; 21np + Sequence Listing; English.
XX	
CC	The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (AB101840-AB116175), expressed DNA sequences (AB101840-AB116175) and the encoded proteins (ABB57737-ABB72072).
CC	The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 1638 AA;

Query Match	8.0%;	Score 163;	DB 22;	Length 1638;
Best Local Similarity	19.8%;	Pred. No. 1.8e-07;		
Matches	85;	Conservative	62;	Mismatches 155;
			Indels	128;
			Gaps	16;

```

QY 27 GEPELECGPNSITIN-----ENTRAEPBGHYVYVGLYDQESCRNDSGROVACISLP- 78
Db 1060 GRSQECEDADVOTCNEBGMETITRPEBGFGRITYTGFDRCFRKNGGTAVNRLSGPQ 1119
QY 79 -FDCSNARTKSLNPRGIEVTTTVISFHPLEFTKVDARVPOCF-----122
Db 1120 GYPCQGTQR-----YEDLTILNVVGFSDNNGVTSBDRKYNLTCIFRGGEAVVSSGIG 1173
QY 123 -----YMEDKTVSAQIEVSEITTAFTOTIVPMPCVCRYEILDGPT-----GQPV 167
Db 1174 AGSSSPIPIEFLPAENLTSKVRLS-----ILYOGPPTTIAVGDPL 1215
QY 168 QFALIGQPVYHKMTCDSEVDTFCAVVHSCFVDNGDGTVEILTMNDGALRYLLNNEX 227
Db 1216 TFRLEADGGINHY-----DIFATNVVA--KDPYSGRSIQULDRFCQCPDVEFVPELDK 1267
QY 228 PTD---IMAGOEAHVYKYADRSOLFYOQOISITIKERNSECVRPOCSEPOGFCAVKTGA 284
Db 1268 LRDDDTLEA--RFNAEFIPESNLFVEA---TVRSRCGQCPAYCQBPAG-----1312
QY 285 AAKRAAAROLRLKKRS--AEPE-----NIIDVRDINLLESDNQALPVD-- 329
Db 1313 RQESFSGRRKRSINTIETPEPELALLESSQLEASTLDEVTYVNVSTVASILGOVPLNET 1372
QY 330 -----LHRAALLOHNGQPVYLAAVQNGISDPSFGSMFGLISA 368
Db 1373 QLGKETEPEEPQVREMIIEVFETREIKESYPRKILVAPVETVCMPT---AEYHGLITA 1429
QY 369 LIAAVITIS 378
Db 1430 IILMLILFS 1439

```

```

XX RESULT 10
XX ABB67239
XX ID ABB67239 standard; Protein; 611 AA.
XX AC ABB67239;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster polypeptide SEQ ID NO 28509.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX   pharmaceutical.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX PA (PEKE ) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EM;
XX WPI: 2001-656860/75.
XX DR N-PSDB; ABL11342.
XX New isolated nucleic acid detection reagent for detecting 1000 or more

```

DR	N-PSDB; ABLI14157.
XX	
PT	New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -
PR	
PP	
XX	
PS	Disclosure; SEQ ID NO 36954; 21pp + Sequence Listing; English.
XX	
CC	The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (AB116176-ABL30511), expressed DNA sequences (AB101840-ABL16175) and the encoded proteins (ABB57737-ABB72072).
CC	The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
CC	
CC	
XX	
SQ	Sequence 611 AA;
	Query Match 8.0%; Score 162.5; DB 22; Length 611;
	Best Local Similarity 21.4%; Pred. No. 4.1e-08;
	Matches 63; Conservative 51; Mismatches 126; Indels 55; Gaps 13.
OY	30 EIECGPSTITTFNFRNAFECHVYVKGLYDDEGCRNDDEGRVAGISLPFD---SCNVA 85
Db	88 QYCCKTMTMRVNIERDFRYGMIFSKRGYSRDNHCYLKRG--TGHSATFELFLNSCGMT 145
OY	86 RTRSRN-----PRGIPTTTVVVISPHRLFVTIKVDRAVRVOC---PYMEADKTVSNO 133
Db	146 SSANINMAGCYGAPTPSGSYVENTIIIOXDYVVOEWDQARKRLCTWYDFYEKAATFERPQ 205
OY	134 IEVSITIA-----EQIQIVMPRCREIINDG-GRTGQPVGFAL-IGRPVYNHKKTCSE 185
Db	206 VDMLEAVTAANFLGDNLQ-----CMNQIDVGKGPWASEVSGVIKIGQTMTMLAIKDD 257
OY	186 TVDFCAVAVHSCFVDGNGDFVEILNADGCALDKYLNNLEYPDTLMAGEAHV----- 239
Db	258 E-NKEDMLVRNCVANHDGRARPIQLVDQNGCVARRKIMSKFOKIKNF--GPSASVYSFAFY 314
OY	240 --YKADBSQLFYOQCOISITTIKEPNSECVRPCSEPGFGAAYKTGAATAKRRAAA 292
Db	315 QAQFKRPDSMNVMHFQCVIDQVC---RYNCBPERKC-----GPELPGGEYGCLPQIGA 359
	RESULT 12
	ABB58337
ID	ABB58337 standard; Protein; 758 AA.
XX	
AC	ABB58337;
XX	
JT	26-MAR-2002 (first entry)
DE	Drosophila melanogaster polypeptide SEQ ID NO 1803.
XX	
KW	Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.
XX	
OS	Drosophila melanogaster.
XX	
MO	WO200171042-A2.
XX	
PJ	27-SEP-2001.
PD	
PF	23-MAR-2001; 2001WO-USO9231.
XX	
PR	23-MAR-2000; 2000US-191637P.
XX	
PA	11-JUL-2000; 2000US-0614150.
XX	
XX	(PEKE) PE CORP NY.
XX	

PI Venter JC, Adams M, Li PWD, Myers EW;
 XX MPI; 2001-656860/75.
 DR N-PSDB; ABL02440.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Disclosure; SEQ ID NO 1803; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL1840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX Sequence 758 AA;
 SQ
 Query Match 7.4%; Score 150; DB 22; Length 758;
 Best Local Similarity 23.6%; Pred. No. 1.3e-06;
 Matches 81; Conservative 50; Mismatches 114; Indels 98; Gaps 20;
 OY 31 IECGPTSIINFTRNAFEGHYVVGKLYDQECGRNDEGGRQVAGISLPD-----SC 82
 DB 393 ICRSGEMITKIRTSKLFQGYAKGA--PKCAVNVN-----SLEDFMKRYNDELC 444
 OY 83 NVART---RSLNPRGIFVTTTVVISFHPLEFVKVDRAYVOCFYMEADKTVSAQIEV--- 136
 DB 445 NVQSAVYGYVM-----DIVQIHDIMIVTSSDLGLAVSCQYDLTNKTVANNVDLGV 496
 OY 137 SETTFAFOFOIY---PMPCRYEILDGPTGQPVQFAIGQVYHKMTCDSFVDT--F 190
 DB 497 GEIESLSLEIITIDSPNVIMKTARDG---SDMKRIAEVGDPLALRF---ETVDANS 549
 OY 191 CAVHSCFVDGNGDPTVEI--LNADGALDKYLLNNLEYPTDLMAGQEAHYKYA-DRSQ 247
 DB 550 ELFVRLVAMDGT-DSAEITLIDANGCPTDQYIMSAMQ-----KLANNRKY 594
 OY 248 LFYQCOISITIKEPNSECVR-----PQCEP-----QGFQAVK----- 280
 DB 595 LLSQED---AFKPSSELYQFALVTPCIPRC-EPVTCNDENGELKSLISGRRRSYL 650
 OY 281 --TGGAARPAQAOLRLKRSAPENIIDVRTDINTLEISD 321
 DB 651 NGTDGVELAIKSEROKRDVSHQAADENILVQ---STQID 689
 RESULT 13
 ABB58185
 ID ABB58185 standard; Protein; 833 AA.
 XX
 AC ABB58185;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 1347.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX MPI; 2001-656860/75.
 DR N-PSDB; ABL02288.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Disclosure; SEQ ID NO 1347; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL1840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX Sequence 833 AA;
 SQ
 Query Match 7.0%; Score 142; DB 22; Length 833;
 Best Local Similarity 22.7%; Pred. No. 1.1e-05;
 Matches 88; Conservative 51; Mismatches 128; Indels 120; Gaps 22;
 OY 31 IECGPTSIINFTRNAFEGHYVVGKLYDQECGRNDEGGRQVAGISLPD-----SC 82
 DB 381 ICRSGEMITKIRTSKLFQGYAKGA--PKCAVNVN-----SLEDFMRYNDELC 432
 OY 83 NVART---RSLNPRGIFVTTTVVISFHPLEFVKVDRAYVOCFYMEADKTVSAQIEV--- 136
 DB 433 NVQSAVYGYVM-----DIVQIHDIMIVTSSDLGLAVSCQYDLTNKTVANNVDLGV 484
 OY 137 SETTFAFOFOIY---PMPCRYEILDGPTGQPVQFAIGQVYHKMTCDSFVDTFCV 193
 DB 485 GEIESLSLEIITIDSPNVIMKTARDG---SDMKRIAEVGDPLALRFEL-VEPNPYLEF 540
 OY 194 VHSCEFVDGNGDPTVEI--LNADGALDKYLLNNLEYPTDLMAGQEAHYKYA-DRSQ 250
 DB 541 VRELVAMDS-DSAEITLIDANGCPTDQYIMGTQ-----KLANNRKY 585
 OY 251 QCOISITIKEPNSECVR-----PQCEPQGFQAVKGTGAARPAQAOLRL----- 297
 DB 586 QPD---AFKPSSELYQFALVTPCIPRC-EPV-----ICDESDASGELKSLVSYG 633
 OY 298 -KKRSAPENIIDVRTDINTLEISDQNALPVDLRRLALQHNQGPVILAAVQNGICMSP 356
 DB 634 RKKRSV-----LN-----GTDGAELISTNHR-----RDGVPV----- 661
 OY 357 GFSGMEGLIALAAVITTSFKFRP 383
 DB 662 -----DDNILLMQSIQITQKFGQP 681
 RESULT 14
 ABB70535
 ID ABB70535 standard; Protein; 692 AA.
 XX
 AC ABB70535;
 XX
 DT 26-MAR-2002 (first entry)
 XX

DE	Drosophila melanogaster polypeptide SEQ ID NO 38397.
XX	
KW	Drosophila; developmental biology; cell signalling; insecticide;
KW	pharmaceutical.
XX	
OS	Drosophila melanogaster.
PN	WO200171042-A2.
PD	
XX	27-SEP-2001.
XX	
PE	23-MAR-2001; 2001WO-US09231.
XX	
PR	23-MAR-2000; 2000US-191637P.
DR	11-JUL-2000; 2000US-0614150.
XX	
PA	(PERKE) PE CORP NY.
PI	Venter JC, Adams M, Li PWD, Myers EW;
XX	
WI	WPI: 2001-656860/75.
DR	N-PSDS; ABL14638.
XX	
PT	New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -
PT	
XX	
PS	Disclosure; SEQ ID NO 38397; 21pp + Sequence Listing; English.
CC	
CC	The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB72072).
CC	(ABB57737-ABB72072).
CC	The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
CC	
XX	
SO	Sequence 692 AA;
	Query Match 6.9%; Score 140; DB 22; Length 692;
	Best Local Similarity 20.8%; Pred. No. 1.4e-05;
	Matches 63; Conservative 59; Mismatches 141; Indels 40; Gaps 16;
OY	31 IECGPTSTINFNNTNAFEGHYVVGKLDQEGCR---NDECGROYAGISLPFGSCNVAR 86
DB	358 VHCKDTRIAVOYRTIKKPNRGRIYAIG--RSEFCNIIDIVNSDFR--LDITMAGODCN--- 410
OY	87 TRSLNPGRIFVTTVVISFHPLELVTKVDRAVYEVOCFYMEADKTVS---AQIEVSETTAF 143
DB	411 TQSVT--GVY-SNTVVLDHSHVMRKADIKYVKCTYDMSSKNITFGMPIMPDPKHIN 467
OY	144 QTQIYPMVVCREYIIIDGPTGPVOFAITGGPVYHKWTCDSETVDTFCVAVHSCFEVDGN 203
DB	468 SSPEAPPP--RIRIID--TROREVTVIRIGRLNFRIEIPEDT--PYGI FARSCVAMAKD 521
OY	204 GDT-VEILNAGCALDKTLNLNLEVPDLMAGEAH-YKYVADRQLFPQCQISTITKEP 261
DB	522 ARTSKRIIDDDGCPPDPTIFPG--TTAAGNALQSTYEAFREFYESYGVIPOCVNKY----- 574
OY	262 NSECVRPCCEPOFGAVKGTGAANKPAAALDLRLTKRSAPENIIDVRTINPLEISD 321
DB	575 ---CLGF--CER---AVCEMNNDSEFSIGRRRRRSIESNDTKSESDMINISOELLVIDFGD 626
OY	322 DNQ 324
DB	627 EKR 629
RESULT	15

[illegible]


```

Dd 312 TDNYGQGVGHEAGSTGAGTYFENITIVIOYDPOVEWMOQARKLCTMHDQYKESVTER-- 369
Qy 136 VSEITTAFOQIVMPV-----CRYEILDG-GPTGQVQVQAI-IGQPVY 177
Dd 370 -----PPVMDLDVBRADPAGDNVGCWMOIQVKGFWASEVSLVXIGQTM 416
Qy 178 HKMTCDSEVDTFCVAVVHSCFVDGNGDVEILNADGALDKYLLN-----NLEYPDL 231
Dd 417 MVLAIKDD--SKFLMLVNCVAHDGKRAPILQVDRGCVTPKMLSRRTKIKNGASASV 475
Qy 232 MAGGEAHVYKADRSQLEFYQCQISITIKERNSEC-----VRPCCS---EPQ 274
Dd 476 LSYAHFOAFKFPDSEVHFOCTIQICRYHCEPCSAETNLODVHHLQVGPESQYGPPL 535
Qy 275 GFGAVTKGGAAKPMAAOLRLKKRSAPD-----NIIDVTDINTLEISDNOA----- 325
Dd 536 HVDVAVHASAIGKRDRERRVRRARAFAEPQVGLNRIIKV-VSSGDLTFALDDQAGNGS 594
Qy 326 -----LPVDLRHRLALQHNQPVILAAVQNG-ICMSPFGFSMFGSLIALIAVILP-- 376
Dd 595 STNGANGVD-----RSPQTMVFPPLREGLICMTTPGFAITLIVLGLIVTSCLTSA 646
Qy 377 -ISFKRP 383
Dd 647 VLYVRLRP 654

RESULT 18
ABB68164
ID ABB68164 standard; Protein; 418 AA.
XX AC ABB68164;
XX Df 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster polypeptide SEQ ID NO 31284.
XX KM Drosophila; developmental biology; cell signalling; insecticide;
XX KM pharmaceutical.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX PA (PEKE ) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX DR WPI: 2001-656860/75.
XX DR N-PSDB; ABL12267.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions -
XX PS Disclosure; SEQ ID NO 31284; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABL16176-ABJ30511), expressed DNA
XX CC sequences (ABL1840-ABL16175) and the encoded proteins
XX CC (ABB57737-ABB72072).

```

```

CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX S0 Sequence 418 AA:
Query Match 5.9%; Score 119.5; DB 22; Length 418;
Best Local Similarity 19.8%; Pred. No. 0.001;
Matches 78; Conservative 67; Mismatches 103; Indels 65; Gaps 16;
Qy 27 GEPEIEGCPITITINE-----NTRNAFEGHYVYKGL--YDQGRND-EGGRQVAGISL- 77
Dd 29 GSHKVCSEDDMRVDIGLPDAESKDSAPQIYLEGLKGPDERCQPOQIDGSLAVFRLSL 88
Qy 78 PEDSCNVARTRSLNPRGIFVTTVYISPHPL-----YTKVBRAYVQCFY 123
Dd 89 DFYECGV--TKMVO-----LTKKVVYHKIITESTSKEIVSKCITIASPAIVMMNA 141
Qy 124 MEADKTVSA-----QIEVSEITTAFOQIVMPVPCRYEILDGPTG 164
Dd 142 TTGSSSTSSGGIHGLYKRDVLPAGPQEPDLETTLSL-TKRAEPRLSIGVSQDGKF 200
Qy 165 QPVQFAITGQPVYHKMTCDSETVDFCAVHSCFVDGNGDVEILNADGALDKYLLNN 224
Dd 201 TRDLTVKSGTFLTMEINDEDSAPYGLGVNLDVTDTHSS-ETLIFKCGTVDPYLPEN 259
Qy 225 LEYPTDLMAQGEAHVYKADRSQLEFYQCQISITIKERNSECVRPCCSEPO-GFGAVKTGG 283
Dd 260 FNTIDGDLISAKFAKFPDSSVVOFRAVNVCL---DKLGRQCSNNQVGFGRKREI 315
Qy 284 AAAPRAAAAOURL-LKRSAPEN---IIVRTDINTLEISDNOALPVDLRHRLALOH- 338
Dd 316 SSANKVYRISLAMPLOQVODIEGVKNKEVLOEKKIREKLA--NORLARNSRGNAEMQT 373
Qy 339 --NGQPVILAAVQNGICMSPFGFSMFGSLIAL 369
Dd 374 PASAQPAFVYDERELGHLISAGSAAASGLSLAL 406

RESULT 19
ABB71434
ID ABB71434 standard; Protein; 2284 AA.
XX AC ABB71434;
XX Df 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster polypeptide SEQ ID NO 41094.
XX KM Drosophila; developmental biology; cell signalling; insecticide;
XX KM pharmaceutical.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX PA (PEKE ) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX DR WPI: 2001-656860/75.
XX DR N-PSDB; ABL15537.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions -

```

```

XX PS Disclosure; SEQ ID NO 41094; 21bp + Sequence Listing; English.
XX CC
CC CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (AB161737-AB172072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ
SQ Sequence 2284 AA;

Query Match
Best Local Similarity 19.9%; Score 111; DB 22; Length 2284;
Matches 61; Conservative 46; Mismatches 97; Indels 102; Gaps 18;

QY 30 EIECGPIS---ITINENTRNAFEGHYVYKGLYDQEGC---RNDEGROVAGISLPDSCN 83
   ||| | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
DB 59 DVKCDQSGMAMVEVEFS--EDFEGVYISQGYFSDPKCNVYKGRSFT-FVVPYDGC- 114
   ||| | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 84 VARTRSLNPRGIFVTYVVISFHPLEFVTKYDR-----AYRQCFY-WEADKTVSAQI 134
   ||| | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
DB 115 -----GSKPSCVCAISIENTILITQDDRIQNSFDIARIISCRGDERKTVYFKP 164
   ||| | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 135 EYSEITTAFOQTQIVPMFV-CRYEILDGGPYG-QPVQPAI-IGOPVYKRWTCDETVDTFC 191
   ||| | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
DB 165 FVVDMLLEVISVDTPSGFVEECMEIGTGTPTPNVAKPIQGLTLGLDI-----TFT 212
   ||| | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 192 A-VVHS-----CFVVDG-----NGDIYEILNADGCAIDKYL-----L 222
   ||| | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
DB 213 INVKHSQADINILQCYASDDMDFEARTTKRLQLSDKRCSTIKETIFGEMRFEAGSSL 272
   ||| | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 223 NMLEYPTDLMAGQEAHYKYADRSOLFQCOISITTIKEPNSGCVR----- 267
   ||| | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
DB 273 TSYTYNT-----LKAFRPDRSQVYLKCDIELC-----NGACKRDYTCGSLNSLCPES 321
   ||| | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 268 ---PQC 270
   |||
DB 322 STDPOC 327

RESULT 20
ABB68660
ID ABB68660 standard; Protein; 768 AA.
XX AC
XX ABB68660;
XX DT
XX 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster polypeptide SEQ ID NO 32772.
XX XX
XX KW Drosophila: developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX OS Drosophila melanogaster.
XX PN
XX WO200171042-A2.
XX PD
XX 27-SEP-2001.
XX PF
XX 23-MAR-2001; 2001WO-US09231.
XX PR
XX 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-0614150.
XX (PEKE ) PE CORP NY.
XX PA
XX Venter JC, Adams M, Li PWD, Myers EW;

```

```

XX DR WPI; 2001-656860/75.
XX DR N-PSDB; AB112763.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions -
XX PS
XX PS Disclosure; SEQ ID NO: 32772; 21bp + Sequence Listing; English.
XX CC
CC CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (AB161737-AB172072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ
SQ Sequence 768 AA;

Query Match
Best Local Similarity 16.3%; Score 100.5; DB 22; Length 768;
Matches 51; Conservative 55; Mismatches 120; Indels 87; Gaps 11;

QY 33 CGPTSTINENTRNAFEGHYVYKGLYDQEGCRNDEGRQY-----AGISLPDSCVART 87
   ||| | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
DB 330 CTNDEMFTIKNPDMFVGKITY-SMHSKDCIARSGNSGSLTLQIGSEVKEKRGGITLRA 388
   ||| | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 88 RSLNP--RGIEVTTTVVISFHPLEFVTKVDRAVYKVCFMHADKTVSAQIEVEITP---- 141
   ||| | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
DB 389 YEMTQYORTFISALVVIQNNPMTQGRILKVGCTQSNATTSIGVSRDSVSSEPV 448
   ||| | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 142 ----AFQQTQIV-----PMPCRIEILD-----GGPIGQPVQ 168
   ||| | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
DB 449 PSALIESSLEYTEHMFPHGVAHYNNSGTGPHHPSISLQILDLSHQHETNDVQIQNDLE 508
   ||| | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 169 FALIQG-----PYVHKWTCSEYVDTFCVAVHSCFVDDGNDPVEILN 211
   ||| | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
DB 509 LQIVAEYSPQQLAEHMELOLAPLPDFRATISLAKTAD-----NENFVLLID 554
   ||| | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 212 ADGCALDKYL---LNNLEYPITDLMAGQEAHYKYADRSOLFQCOISITTIKEPN-SECV- 266
   ||| | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
DB 555 ERGCPTDASVPFALERVHTASRSMLRARFHAFFSGTAVNSVDVAKIRFVENCSPSNCIS 614
   ||| | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 267 -----RPQCSBP 273
   |||
DB 615 SSMORRRROADP 627

RESULT 21
AAR20112
ID AAR20112 standard; Protein; 935 AA.
XX AC
XX AAR20112;
XX DT
XX 06-APR-1992 (first entry)
XX DE AE-III (peptidylhydroxyglycine N-C lyase precursor).
XX XX
XX KW Amidation; PHL.
XX OS Xenopus laevis.
XX PN
XX Key Location/Qualifiers
XX FH 1..935
XX FT Protein
XX FT /label= AE-III
XX FT /note= "including PAM and PHL domains"
XX FT 383..935
XX FT /*tag= b

```

```

FT      /Product= PHL
XX      EP465404-A.
XX      08-JAN-1992.
XX      27-MAY-1991; 91EP-0810399.
XX      30-NOV-1990; 90JP-0329911.
XX      01-JUN-1990; 90JP-0141678.
XX      10-AUG-1990; 90JP-0210535.
XX      (CIBA ) CIBA GEIGY AG.
XX      Iwasaki Y, Shimoi H, Suzuki K, Ghisalba D, Nishikawa Y,
PI      Kawahara T, Kangawa K;
XX      WPI: 1992-010570/02.
XX      N-PSDB; AAQ20269.
XX      PT Novel DNA encoding peptidyl hydroxy:glycine N-C lyase (PHL) -
XX      PT used to prepare PHL which can be used in the amidation of
XX      peptide(s) e.g. human calcitonin.
XX      PS Claim 4; Page 18; 28pp; English.
XX      CC The sequence was deduced from a cDNA insert from pAE-III-202-4
XX      CC (FERM Bp-3172). The vector serves as a source for a DNA fragment
XX      CC encoding PHL for the construction of an expression vector for the
XX      CC prep. of recombinant PHL. The protein may be truncated to
XX      CC comprise only residues 383-706 or 383-713. The PHL catalyses the
XX      CC reaction: R-GlyOH -> R-NH2. It can be used to produce peptides with
XX      CC amidated C-termini, e.g. calcitonin, growth hormone, LH-RH.
XX      SQ Sequence 935 AA;

Query Match      4.9%; Score 100; DB 13; Length 935;
Best Local Similarity 19.4%; Pred. No. 0.47;
Matches 76; Conservative 57; Mismatches 148; Indels 110; Gaps 18;

OY      33 CGPTSTITINFNTRNAEGHYVYKGLYDQEGCRNDEGGROYAGISLPDSCNVARTRSLNP 92
DB      526 CQPTDVAVDPRITGNFVADGY-----CN-SRIQFSP 556
OY      93 RGFV-----TTVYISFHPLEVTYKVDRAVRCFTYMEADKTVS 131
DB      557 NGMFIQMGSEITSSNVRPGQFRIPHSITMVPDQGOGLCVADRENG-RIDCFHAETGNFYK 615
OY      132 AQEVESEI-TTAFQTQIVMPVCRYEILDGGPTGPOFAIIGOPYVHKMTCDSERYDTF 190
DB      616 -QIKHGFEGREYVAVSYAP-----GG-----VLAVNGKPYT-----GY 648
OY      191 CAVVHSCFYVDGNGDTVEILMDGALDKYL--LNNLEYPTDLMAGQEAHYVRYADRSOL 248
DB      649 SAVYQGFMLNFSNGD-----ILDFPIARKNFDMDHDAADDDGVYVGDAAHANA 698
OY      249 FVQCQS-----ITIKEPN--SECVRPQCEPFGFCAVGTGGAAPRAAAQRLKRS 301
DB      699 VMKFSKAEHRSVKAGIEVEITETETETETIRSRPKNESVEKOTQKQOK--QKNS 756
OY      302 A-----EPENII-DVRDTINTLLE---ISDNOALPYDLRRHALLOHNGOPYVILAAVONGI 352
DB      757 AGVSTQEKONVVOEINAGVPTQKONVVOESSAGVSTQEKQVVOSSAGVSTQEKQSVV 816
OY      353 CMSPFGEFSMFMGISIALI-AAVITITISFKFR 382
DB      817 QESSAGVSFVLITLITLIPAVLIAIAIFIR 847

RESULT 22
ABB71814
XX      ABB71814 standard; Protein; 227 AA.

```

```

AC      ABB71814;
XX      DT 26-MAR-2002 (first entry)
XX      DE Drosophila melanogaster polypeptide SEQ ID NO 42234.
XX      DE Drosophila melanogaster polypeptide SEQ ID NO 42234.
XX      KM Drosophila: developmental biology; cell signalling; insecticide;
XX      KM pharmaceutical.
XX      OS Drosophila melanogaster.
XX      PN WO200171042-A2.
XX      PD 27-SEP-2001.
XX      PF 23-MAR-2001; 2001WO-US09231.
XX      PR 23-MAR-2000; 2000US-191637P.
XX      PR 11-JUL-2000; 2000US-0614150.
XX      PA (PEKE ) PE CORP NY.
XX      PI Venter JC, Adams M, Li PMD, Myers EW;
XX      DR WPI: 2001-656860/75.
XX      DR N-PSDB; ABL15917.
XX      PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX      PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX      PT interactions -
XX      PS Disclosure; SEQ ID NO 42234; 21pp + Sequence Listing; English.
XX      CC The invention relates to an isolated nucleic acid detection reagent
XX      CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX      CC useful in developmental biology and in elucidating cell signalling and
XX      CC cell-cell interactions in higher eukaryotes for the development of
XX      CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX      CC discloses genomic DNA sequences (AB101840-AB116175), expressed DNA
XX      CC sequences (AB101840-AB116175) and the encoded proteins
XX      CC (AB157737-AB172072).
XX      CC The sequence data for this patent did not form part of the printed
XX      CC specification, but was obtained in electronic format directly from WIPO
XX      CC at ftp.wipo.int/pub/published_pcl-sequences.
XX      SQ Sequence 227 AA;

Query Match      4.8%; Score 97; DB 22; Length 227;
Best Local Similarity 22.6%; Pred. No. 0.1;
Matches 30; Conservative 29; Mismatches 52; Indels 22; Gaps 6;

OY      22 DNGVEGEPEIECGPTSTITINFNTRNAEGHYVYKGLYDQEGC-----RNDEGGROYAGI 75
DB      46 DQGIQ-KVNLKCGADSNVVLLETEKPMGYMTRGSPYKXSAFCPMKSSQGSRTME-M 103
OY      76 SLPPDSCNVARTRSLNPGRGIFVTTTVYISFHPLEVTYKVDRAVRCFTYMEADKTV 130
DB      104 NFQUDQQTIRD-----GDLTYNIIVYIQNDPELITPGDSAFLECDFRQPRNLDVQASM 157
OY      131 SAQIEV---SEIT 140
DB      158 QARDRVATGSKIT 170

RESULT 23
AAP94856
XX      AAP94856 standard; protein; 693 AA.
XX      AC AAP94856;
XX      DT 27-JUN-1990 (first entry)
XX      DE Expression plasmid pUCP1C1799 BglII gene product.

```

```

XX alpha-amidating; pAX799; alpha amide; ds.
XX Synthetic.
XX EP299790-A.
XX 18-JAN-1989.
XX 15-JUL-1988; 88EP-0306508.
XX 17-JUL-1987; 87JP-0177184.
XX 05-DEC-1987; 87JP-0306867.
XX (SUNR) SUNTORI LTD.
XX Ohsuye K, Kitano K, Tanaka S, Matsuo H, Mizuno K;
XX MPI: 1989-017279/03.
XX N-PSDB: AAN94527.
XX WPI: 1989-017279/03.
XX N-PSDB: AAN94527.
XX Recombinant C terminal alpha amidating enzymes of Xenopus laevis -
XX and their precursors deoxyribonucleic acid encoding sequences.
XX Disclosure; 107pp; English.
XX PS The sequence encodes a derivative of the mature C-terminal alpha-
XX amidating enzyme from plasmid pXA457.
XX CC The plasmid was screened from an E.coli library using plasmid pXA457
XX to screen a larger library.
XX CC Although pXA799 is similar to pXA457 at the N-terminus, it has an area
XX of hydrophobic elements suggesting a membrane function.
XX See also AAN93060.
XX CC
XX SQ Sequence 693 AA;

Query Match 4.7%; Score 94.5; DB 10; Length 693;
Best Local Similarity 21.6%; Pred. No. 1.1;
Matches 50; Conservative 25; Mismatches 66; Indels 91; Gaps 12;

QY 33 CGPTTITNFNTRNAFECHVYKGLYDQGCNDEGGRQVAGISLPDSCNVAARTSLNP 92
DB 491 CQPTDVAADPTGNGFVADG-----CN-SRIMQSP 521
QY 93 RGEIV-----TTTVISFHPLEVTKYDRAYRVQCFYMEADKTVS 131
DB 522 NCMFTMONGEETSNLPRGQFRIPHSITMISDQGLCVADRENG-RIOCFHAKTGEFVK 580
QY 132 AQIEVSEI-TTAFQIQIYMPYCRREILDGPTGQPVQFALIGQVYHKMTCDSETVDTF 190
DB 581 -QIKHQEFGREYFAVSAP-----GG-----VLYAVNGKPYG---DSTPVQGF 620
QY 191 CAVVSCFVDGNGDPTVEILNADGALDKYL--LNMLEYPTDLMAGQEAHY 240
DB 621 -----MLNFSNGD-----ILDFTIPARKNFEKPHDIAGDDGTIV 655

RESULT 24
AAP94854
ID AAP94854 standard; protein; 875 AA.
XX
XX AAP94854;
XX
XX 27-JUN-1990 (first entry)
XX
XX C-terminal prepro-C-terminal alpha-amidating enzyme of pXA799.
XX
XX alpha-amidating; pAX799; alpha amide; ds.
XX Synthetic.
XX EP299790-A.

```

```

PD 18-JAN-1989.
XX
XX 15-JUL-1988; 88EP-0306508.
XX
XX 17-JUL-1987; 87JP-0177184.
XX 05-DEC-1987; 87JP-0306867.
XX (SUNR) SUNTORI LTD.
XX Ohsuye K, Kitano K, Tanaka S, Matsuo H, Mizuno K;
XX MPI: 1989-017279/03.
XX N-PSDB: AAN90791.
XX WPI: 1989-017279/03.
XX N-PSDB: AAN90791.
XX Recombinant C terminal alpha amidating enzymes of Xenopus laevis -
XX and their precursors deoxyribonucleic acid encoding sequences.
XX Disclosure; 107pp; English.
XX PS Plasmid pXA799 contains a sequence derived from Xenopus laevis.
XX CC The plasmid was screened from an E.coli library using plasmid pXA457
XX to screen a larger library.
XX CC Although the gene product is similar to that of pXA457 at the
XX N-terminus, it has an area of hydrophobic elements suggesting a membrane
XX function. See also AAN93060.
XX CC
XX SQ Sequence 875 AA;

Query Match 4.7%; Score 94.5; DB 10; Length 875;
Best Local Similarity 21.6%; Pred. No. 1.6;
Matches 50; Conservative 25; Mismatches 66; Indels 91; Gaps 12;

QY 33 CGPTTITNFNTRNAFECHVYKGLYDQGCNDEGGRQVAGISLPDSCNVAARTSLNP 92
DB 530 CQPTDVAADPTGNGFVADG-----CN-SRIMQSP 560
QY 93 RGEIV-----TTTVISFHPLEVTKYDRAYRVQCFYMEADKTVS 131
DB 561 NCMFTMONGEETSNLPRGQFRIPHSITMISDQGLCVADRENG-RIOCFHAKTGEFVK 619
QY 132 AQIEVSEI-TTAFQIQIYMPYCRREILDGPTGQPVQFALIGQVYHKMTCDSETVDTF 190
DB 620 -QIKHQEFGREYFAVSAP-----GG-----VLYAVNGKPYG---DSTPVQGF 659
QY 191 CAVVSCFVDGNGDPTVEILNADGALDKYL--LNMLEYPTDLMAGQEAHY 240
DB 660 -----MLNFSNGD-----ILDFTIPARKNFEKPHDIAGDDGTIV 694

RESULT 25
AAR73053
ID AAR73053 standard; protein; 875 AA.
XX
XX AAR73053;
XX
XX 06-NOV-1995 (first entry)
XX
XX Peptidyl C-terminal alpha-amidating enzyme.
XX DE Peptidyl C-terminal alpha-amidating enzyme.
XX KW Peptidyl C-terminal alpha-amidating enzyme; AE; cell culture;
XX KW Trichostatin; CHO.
XX Not specified.
XX
XX Key Location/Qualifiers
XX FT Peptide 1..39
XX FT /label= Sig_peptide
XX
XX EP649900-A.
XX
XX 26-APR-1995.
XX
XX 07-SEP-1994; 94EP-0306587.

```


[illegible]

XX PS Claim 6; Page 113-114; 122pp; English.

XX CC The present sequence is the fusion protein construct, pIL2Trim3. This

CC construct comprises of the leader peptide from human interleukin 2 (hIL2)

CC and soluble CD39 (solCD39) protein region, having apyrase activity.

CC Soluble CD39 is constructed by removing the N- and C-terminal

CC transmembrane domains. It retains the capacity to metabolise ADP and ADP

CC at relevant concentrations and the ability to block and reverse

CC ADP-induced platelet activation and recruitment, including platelet

CC aggregation. Soluble CD39 polypeptides are useful for inhibiting

CC angiogenesis. It is useful for the treatment of unstable angina, stroke,

CC myocardial infarction, coronary artery disease or injury, embolism,

CC atherosclerosis, peripheral vascular occlusion, preclampsia, platelet-

CC associated ischaemic disorders including lung, coronary and cerebral

CC ischaemia, thrombotic disorders including coronary, peripheral and

CC cerebral artery thrombosis, intracardiac and venous thrombosis,

CC coagulopathy, deep venous thrombosis (DVT), pulmonary embolism (PE) and

CC transient ischaemic attack. Soluble CD39 is also useful for preventing

CC thrombus formation or reformation, occlusion, reocclusion, stenosis or

CC restenosis of blood vessels or stroke.

XX SQ Sequence 474 AA:

Query Match 4.3%; Score 87.5; DB 21; Length 474;

Best Local Similarity 20.9%; Pred. No. 3.5;

Matches 55; Conservative 40; Mismatches 75; Indels 93; Gaps 13;

QY 113 VDRAYVOCFYME-ADKTYSAQIEVEITTAPOTVMPVCRYEILDGGPGQPVQFAI 171

DB 5 IDRMQLSLCALSLALVYNSSTKTKQLTSSYONKALPENVKYGVILDGGSHTSLYI-- 62

QY 172 IQGPVHKWTCDSFVDFCAVH-----SCFVDDGN----- 203

DB 63 -----YKMPAEKEN-DT--GVVHQVEECRVKGPISKRVQKNEIGIYLDGMEARREY 113

QY 204 -----GDT--VEILMAGCALDKYLLNLE-----YPTD-----LMAGQEAHV 238

DB 114 PRSQHETPVYLGATAGRRLRLMESEELADRVLDVERSLSNYPDPFGARITGQEGE 173

QY 239 VY-----KYADRSOLFYOQISITIKPEPNSCVAPQCSPEPGFCAVNTGGAAPR 288

DB 174 AVGWITINVLKRFSGKTRWF-----STVPEYETNNQ-----ETFGALDGGASTGV 219

QY 289 AAAAQLRLKRSAPENIIDVR 311

DB 220 FVPQNGTIE-----SPDNALQFR 238

RESULT 31

ID AAY70901 standard; Protein; 473 AA.

XX AAY70901;

DT 17-AUG-2000 (first entry)

XX Protein encoded by Trim 4 construct.

XX Soluble CD39; ADP-induced platelet activation; platelet aggregation;

KW unstable angina; myocardial infarction; stroke; coronary artery disease;

KW atherosclerosis; peripheral vascular occlusion; preclampsia; embolism;

KW platelet-associated ischaemic disorder; lung ischaemia; thrombolytic;

KW cerebral ischaemia; thrombotic disorder; coronary artery thrombosis;

KW cerebral artery thrombosis; intracardiac thrombosis; venous thrombosis;

KW peripheral artery thrombosis; coagulopathy; deep venous thrombosis;

KW PE; pulmonary embolism; transient ischaemic attack; thrombus formation;

KW occlusion; reocclusion; stenosis; restenosis; antiangiinal; cardiac;

KW cerebroprotective; antiatherosclerotic; vasotropic; anticoagulant;

KW coronary ischaemia; vascular occlusion; pIL2Trim4 variant.

OS Homo sapiens.

OS Synthetic.

XX FH Key Location/Qualifiers

XX FT Cleavage-site 26..27

XX FT Protein 35..473

XX FT note="Soluble portion of CD39"

PN MO200023094-A2.

XX 27-APR-2000.

XX 13-OCT-1999; 99MO-US23641.

XX 16-OCT-1998; 98US-0104585.

XX 06-NOV-1998; 98US-0107466.

XX 13-AUG-1999; 98US-0149010.

XX (IMV) IMMUNEX CORP.

XX (CORR) CORNELL RES FOUND INC.

XX Maliszewski CR, Gayle RB, Marcus AJ;

XX WPI; 2000-339518/29.

XX Inhibiting platelet activation and recruitment, useful for treating a

XX mammal suffering from unstable angina, myocardial infarction, stroke,

XX coronary artery disease or injury, comprises administering soluble CD39

XX polypeptides -

XX Example 11; Page 114-116; 118pp; English.

XX The present sequence is the protein encoded by Trim4 construct. pIL2Trim4

XX variant was constructed by removing the human IL2 residues from solCD39

XX fusion construct. Fusion of 12 amino acids from the N-terminus of mature

XX human IL2 to the solCD39 coding region results in high levels of both

XX expression and activity in the supernatants of transfected cells. SolCD39

XX is used in the treatment of unstable angina, myocardial infarction,

XX stroke, coronary artery disease or injury, atherosclerosis, peripheral

XX vascular occlusion, preclampsia, embolism, platelet-associated ischaemic

XX disorder including lung ischaemia, coronary ischaemia and cerebral

XX ischaemia, a thrombotic disorder including coronary artery thrombosis,

XX cerebral artery thrombosis, intracardiac thrombosis, peripheral artery

XX thrombosis, venous thrombosis, thrombosis, coagulopathy, deep venous

XX thrombosis (DVT), pulmonary embolism (PE), transient ischaemic attack.

XX Soluble CD39 is also useful for preventing thrombus formation or

XX reformation, occlusion, reocclusion, stenosis or restenosis of blood

XX vessels or stroke.

XX SQ Sequence 473 AA:

Query Match 4.3%; Score 87; DB 21; Length 473;

Best Local Similarity 19.5%; Pred. No. 4;

Matches 51; Conservative 42; Mismatches 77; Indels 92; Gaps 12;

QY 113 VDRAYVOCFYMEADKTYSAQIEVEITTAPOTVMPVCRYEILDGGPGQPVQFAI 172

DB 5 IDRMQLSLCALSLALVYNSSTKTKQLTSSYONKALPENVKYGVILDGGSHTSLYI-- 61

QY 173 GGPVHKWTCDSFVDFCAVH-----SCFVDDGN----- 203

DB 62 -----YKMPAEKEN-DT--GVVHQVEECRVKGPISKRVQKNEIGIYLDGMEARREY 113

QY 204 -----GDT--VEILMAGCALDKYLLNLE-----YPTD-----LMAGQEAHV 239

DB 114 PRSQHETPVYLGATAGRRLRLMESEELADRVLDVERSLSNYPDPFGARITGQEGE 173

QY 240 Y-----KYADRSOLFYOQISITIKPEPNSCVAPQCSPEPGFCAVNTGGAAPR 289

DB 174 XGWITINVLKRFSGKTRWF-----STVPEYETNNQ-----ETFGALDGGASTGV 219

QY 290 AAAAQLRLKRSAPENIIDVR 311

DB 220 FVPQNGTIE-----SPDNALQFR 237

Seq	Sequence	473 AA;	4.38;	Score 87;	DB 21;	Length 473;
Query Match	Best Local Similarity	19.58;	Pred. No. 4;			
Matches	51; Conservative	42;	Mismatches	77;	Indels	92; Gaps 12;
OY	113 VDRAYRVOCFEMADKTVASQAEVSEITRTAROTQIVPVCARYEHLIDGGPTGGPQFAII	172				
DB	5 IDRMOLLSCIALSIALVTVNSSTKTKQLTSSFNKRALPENVYXGIVLDGSSHTSLYT---	61				
OY	173 GQPVHKKWTCDSFETVDFECAVH-----SCEVDGDN-----	203				
DB	62 -----YKMPAEKEN-DT--GVYHVEEBCRVKGPGLSKFVQKVNELGIIYITDOMEAREVI	113				
OY	204 -----GDT--VEILNADGALDKYLLNNLE-----YPTD-----LMAGOEAHV	239				
DB	114 PRSOHETPVYIATAGACRMLRMSEELADRLDIVERSLSNYPDPFGARHITGOEEGA	173				
OY	240 Y-----KYADRSGLFYOCQISITIKRPNSCEVAPQCSSEPGFPAVKTGGAAKPA	289				
DB	174 YGWITINVLGKFSOKTFMF-----SIVPEYTNNO-----ETFGALDGGASTGYT	219				
OY	290 AAOLRLIKRSAPENIIDVR	311				
DB	220 FVPONQITE----SPDNALQFR	237				
RESULT 33						
AAR741171						
ID	AAR741171 standard; Protein: 3038 AA.					
AC	AAR741171;					
DT	18-JAN-1996 (first entry)					
DE	Aspergillus terreus triol polyketide synthase.					
KW	Triol polyketide synthase; TPKS; HMG-CoA reductase inhibitor;					
KM	hypercholesterolaemia; LDL-cholesterol.					
OS	Aspergillus terreus.					
XX						
XX	Key	Location/Qualifiers				
FT	Region	181				
FT	Region	/label= keto-acyl synthase motif				
FT	Region	654..658				
FT	Region	/label= acetyl/malonyl transferase motif				
FT	Region	985..994				
FT	Region	/label= dehydratase motif				
FT	Region	1446..1450				
FT	Region	/label= methyl transferase motif				
FT	Region	1932..1937				
FT	Region	/label= enoyl reductase motif				
FT	Region	2164..2169				
FT	Region	/label= keto reductase motif				
FT	Region	2498				
FT	Region	/label= acyl carrier protein motif				
FT	Misc-difference	282..288				
FT	Misc-difference	/label= misc feature				
FT	Misc-difference	1450..1460				
FT	Misc-difference	/label= misc feature				
FT	Misc-difference	1603..1612				
FT	Misc-difference	/label= misc feature				
FT	Misc-difference	2521..2535				
FT	Misc-difference	/label= misc feature				
XX						
XX	WO9512661-A.					
XX						
XX	11-MAY-1995.					
XX						
XX	28-OCT-1994;	94WO-US12423.				
XX						
XX	02-NOV-1993;	93US-0148132.				
XX						

```

XX (MERI ) MERCK & CO INC.
PA
XX
PI Conder MJ, Davis CR, Hendrickson LE, Mcada PC, Rambosek J;
PI Reeves CD, VincI VA;
XX
DR WPI; 1995-193816/25.
XX N-PSDB; AA092323.
XX
PT Novel DNA encoding triol poly-ketide synthase - used to isolate and
PT identify homologues of triol poly-ketide synthase, and in the treatment
PT of hyper-cholesterolaemia
XX
PS Claim 12; Figure 2; 107pp; English.
XX
XX The full-length TPKS-encoding DNA in plasmid pLOA was
CC designated pPKS100. Splicing of the introns from the DNA
CC sequence and translation of the 914 nt ORF results in a
CC protein of 3038 AAs (AA074171) with a mol. wt. of 269,090
CC daltons. Inspection of the TPKS AA sequence for active
CC site residues and motifs known to be associated with
CC polyketide synthases and fatty acid synthase (FAS) activities
CC resulted in the identification of candidates for expected
CC sites (see FT). Except for the presence of a methyl transferase,
CC not present in FAS, the succession of activities on the
CC TPKS protein is the same as that observed for the rat FAS
CC protein.
XX
SQ Sequence 3038 AA;
XX
Query Match 4.3%; Score 87; DB 16; Length 3038;
Best Local Similarity 19.7%; Pred. No. 77;
Matches 93; Conservative 55; Mismatches 138; Indels 186; Gaps 24;
QY 6 IACCTLLALSYSPVDNGVEPEIEC--GPTSI-----TINFT 44
DB 1175 VAQQTIGA--YSSP-----GDRRLCLVYPHYVDRIPLSLCLATAESGCEKAFMT 1226
QY 45 RNFECHVYVYK---LYDOEGCNDGSGNOVAGISL-PPDSCVATRSILNPGIR--V7 98
DB 1227 INTYDKGDIYLSGDIYVFDAL---QTTLQOVENITTKPSPASPDHAFAMWSGPIIT 1282
QY 99 TTVVVSFHPLEVTYKDR-----AYRQCFYMEADKTVSAQIEVSEITTA---FQTOIVPM 150
DB 1283 PSDLNDPEWMATAQDKKAIPITERIVYPI---RSFLQGLTEERQAQAFHIOKIEML 1339
QY 151 PVCRYEIIIDGPTGPGPYQAIIGQPYVHKWTCSETVDTFCAVVSCFVDD----- 201
DB 1340 E---OVLASANEGR-----HLMYDPGWMENDTEAOIEHLCTANSYHPHRLVQ 1383
QY 202 -----GNGDVEILNDGALDYLLNN-----LEPPDIDMAGOEAHYKXAD 244
DB 1384 RVGQHLLPVRBSNGNFEDLLDHG-LITFYTNTLSFGALHYARLVA-QIAHRYQSM 1441
QY 245 RSQLFYQCOISITIKRPNSECVRPOCSEPOGFAVGTGAARPAQAOL----- 294
DB 1442 ILEI-----GAG---TGATYVYVATPQLGNSITYYDI 1472
QY 295 -----RLKRSAPENIIDVRDINTLEISDNOALP 327
DB 1473 STGFEEQAREQAPFEDRWVFEPLDIRSPADGEP-HAYDLIIASNVLAHTPD---LE 1528
QY 328 VDLRH-RALLONNGPVIL-----AAVONGICMSPP 357
DB 1529 KTMHAHRSILKPGQGVILLETIKHEHTRGLTGLFADMMAGVDDGRCTEPF 1580

```

```

RESULT 34
AAB94352
ID AAB94352 standard; Protein; 636 AA.
XX
AC AAB94352;
XX

```

```

DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:14870.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PS Claim 8; SEQ ID 14870; 2537bp + CD ROW; English.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
XX
XX Claim 8; SEQ ID 14870; 2537bp + CD ROW; English.
XX
XX The present invention describes primer sets for synthesizing 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in
XX the specification. The primer sets can be used in antisense therapy and
XX in gene therapy. The primers are also useful for synthesizing polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
XX AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
XX represent oligonucleotides, all of which are used in the exemplification
XX of the present invention.
XX
SQ Sequence 636 AA;
XX
Query Match 4.3%; Score 86.5; DB 22; Length 636;
Best Local Similarity 21.9%; Pred. No. 7.3;
Matches 84; Conservative 54; Mismatches 153; Indels 93; Gaps 23;
QY 6 IACCTLLALSYSPVD-NGVEGEPEIECGTSTITINNTNNAFEGHYVVGVLQDECR 64
DB 259 VGFLSSLLPQSKSPSRKSPAPGPPQ---PQSSA---KRESFGQ-GTGKDKPTSGAK 309
QY 65 NDEGROYAGISLPDSCNVAATRSILNPG-----IEVTVVVSFHPLEVTYKDRARV 119
DB 310 --DGKSLSLGATGESGSGHQRRRLDQHKERKELEPSTT-----S 349
QY 120 QCFTEADKTVSAQIEVSEITTAFTQTOIVPMPCRYEIIIDGPTG---QPYQ-FAIIGP 175

```

```
Db 350 QC-----AEKKPEASGPEAEPCPELHTEPVE-PLTRAS--SAGPEGGVREPOPIVLGOE 402
Oy 176 VYHKWTCSETVDPFCVAVHSCFVDGNGDVEILNADG-----CALDKY 220
Db 403 EYG-----EHHSSIMH-CRY-DCSGRRVASLDVDGYIKWSENPIMOTKASSISKS 451
Oy 221 LNNLEYPTD---LMAGQEAHVYKYADRSOLFQCOISITIKEPNSECVRPQSEPOGF 276
Db 452 PLTSLIEWATKRDRLLLLGSGVGYRLYDTEAKKMLCEININDMP--RIISLACS-PNGA 508
Oy 277 GAVKTGGAAPAAQAOLRL-----LKKRSAPENIT--DVRTDINTLEISDNOALPV 328
Db 509 SFV---CSAAPSLTSQVDESAPDIGSKMNOVGRILLMTKTKMKOQLGFSLDPEPIAI 565
Oy 329 DLRRHALLQHNGOPVITLAAYONGI 352
Db 566 NC---TAFNHNGNLVLTGAADGVI 586

RESULT 35
AAM40070
ID AAM40070 standard; Protein; 747. AA.
AC AAM40070;
XX
XX
DT 22-OCT-2001 (first entry)
DE Human polypeptide SEQ ID NO 3215.
XX
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
XX leukemia.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
PD 26-JUL-2001.
XX
XX 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
XX (HYSE-) HYSEQ INC.
XX
PA Tang YT, Iau C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou F, Goodrich R, Drmanac RT;
XX
XX MPI; 2001-442253/47.
XX
XX N-SPDB; AAI59226.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
XX PT such as central nervous system injuries -
XX
XX Example 5; SEQ ID NO 3215; 10078bp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and
XX CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
XX CC immunosuppressant and cytostatic activity. The polynucleotides are useful
XX CC in gene therapy. A composition containing a polypeptide or polynucleotide
XX CC of the invention may be used to treat diseases of the peripheral nervous
XX CC system, such as peripheral nervous injuries, peripheral neuropathy and
```

```
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
XX SQ Sequence 747 AA;
XX
XX Query Match 4.3%; Score 86.5; DB 22; Length 747;
XX Best Local Similarity 21.9%; Pred. No.9.4;
XX Matches 84; Conservative 54; Mismatches 153; Indels 93; Gaps 23;
Oy 6 IAFCTTLTALSYSIPVD-NGVEGEPEIECPPTSTITINFNTNNAFEGHYVYGLYDOEGCR 64
Db 259 VGFLSSILLPOSKKSPSRSLSPAGPPO---PQSSA---KRESFGQ-CTGGRDPTSGAK 309
Oy 65 NDEGROYAGISLPFDSQNVARTSLNDRG-----IFVTTTVISFHPLEVTKYDRAVRY 119
Db 310 --DGKSLSLGLATGSGSGWQRORRLQDHGKERKELESTTT-----S 349
Oy 120 QCFYMEADKTVSAQIEVEITTAFOFOIVPMPVCRIYELDGGPVG---QPVQ-FAIIGOP 175
Db 350 QC-----AEKKPEASGPEAEPCPELHTEPVE-PLTRAS--SAGPEGGVREPOPIVLGOE 402
Oy 176 VYHKWTCSETVDPFCVAVHSCFVDGNGDVEILNADG-----CALDKY 220
Db 403 EYG-----EHHSSIMH-CRY-DCSGRRVASLDVDGYIKWSENPIMOTKASSISKS 451
Oy 221 LNNLEYPTD---LMAGQEAHVYKYADRSOLFQCOISITIKEPNSECVRPQSEPOGF 276
Db 452 PLTSLIEWATKRDRLLLLGSGVGYRLYDTEAKKMLCEININDMP--RIISLACS-PNGA 508
Oy 277 GAVKTGGAAPAAQAOLRL-----LKKRSAPENIT--DVRTDINTLEISDNOALPV 328
Db 509 SFV---CSAAPSLTSQVDESAPDIGSKMNOVGRILLMTKTKMKOQLGFSLDPEPIAI 565
Oy 329 DLRRHALLQHNGOPVITLAAYONGI 352
Db 566 NC---TAFNHNGNLVLTGAADGVI 586

RESULT 36
AAB93124
ID AAB93124 standard; Protein; 747. AA.
AC AAB93124;
XX
XX
DT 26-JUN-2001 (first entry)
DE Human protein sequence SEQ ID NO:12006.
XX
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
XX Homo sapiens.
XX
XX EPI074617-A2.
XX
XX 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-0116126.
XX
XX 29-JUL-1999; 99JP-0248036.
XX 27-AUG-1999; 99JP-0300253.
XX 11-JAN-2000; 2000JP-0118776.
XX 02-MAY-2000; 2000JP-0183767.
XX 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
PA
```

```
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
XX Claim 8; SEQ ID 12006; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-ol primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any special methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB93893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
XX Sequence 747 AA:
SQ
Query Match 4.3%; Score 86.5; DB 22; Length 747;
Best Local Similarity 21.9%; Pred. No. 9.4;
Matches 84; Conservative 54; Mismatches 153; Indels 93; Gaps 23;
QY 6 IACCTTIALSYSLPVD-NQVEGEPELEGPTSTTTFNTRNAPEGHVYKGLXDQGR 64
DB 259 VGLSSLPLPSKSPSPRLSPAGGPQ---PQSSA---KESFGQ-GTKGRDPTSGAK 309
QY 65 NDEGROVAGISLPDSCNVARTRSLNPRG---IFVTTTVISFHPLEVTKYDRAVRY 119
DB 310 --DGKSLSLGATGESGWSGHRRLQDHRKEKELEFSTT-----S 349
QY 120 QCFYMEADKTVSAQIEVSEITTAFOQIYMPYCRYEILDGPTG---QPVQ-FAITGQP 175
DB 350 QC---AEKRPEASGPEAEPCPELHTEPVE-PLTRAS--SAGEGGGVREQGFYVLQGE 402
QY 176 VYHKWTCDSVTDTFCVAVHSCFVDGNGDVTVELNADG-----CALDXY 220
DB 403 EYV-----EHSISM-CRV-DCSGRRVASLDVDGKYKWSFNPIQTKASSISKS 451
QY 221 LNNLEYPD---LMAQEAHYKYADRSQLEFYCOISITTEKPEPSECRCPCSPQGE 276
DB 452 PLISLEMAKRRDLILLGSGVGVRLYDTEAKNLCENINDMP--RIISLACS-PNGA 508
QY 277 GAVKTGGAAPAAAOQLRL-----LKRSAPEENII--DVRTDINTLEISDDNOLAVY 328
DB 509 SFV---CSAAPFSLTQVDFASPDIGSKGNQVGRLLMDTKMQQLQFSLDPPPIAI 565
QY 329 DLHRRLDLQHNQPVILAVONKI 352
DB 566 NC---TAFNHNGMLVTGADGVI 586
```

RESULT 37
ABP38294

```
ID ABP38294 standard; Protein; 438 AA.
XX
XX AC ABP38294;
XX
XX 24-JUL-2002 (first entry)
XX
XX DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:3139.
XX
XX KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
XX antibacterial; gene therapy.
XX
XX OS Staphylococcus epidermidis.
XX
XX PN US6380370-B1.
XX
XX PD 30-APR-2002.
XX
XX PF 13-AUG-1998; 98US-0134001.
XX
XX PR 14-AUG-1997; 97US-055779P.
XX
XX PR 08-NOV-1997; 97US-064964P.
XX
XX PA (GENO-) GENOME THERAPEUTICS CORP.
XX
XX PI Doucette-Stamm LA, Bush D;
XX
XX DR WPI; 2002-381255/41.
XX
XX DR N-PSDB; ABN90839.
XX
XX PT Novel isolated nucleic acid encoding a Staphylococcus epidermidis
XX polypeptide, useful for diagnosing and treating bacterial infections -
XX disclosure; SEQ ID 3139; 267pp; English.
XX
XX PS
XX
XX CC ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
XX frame (ORF) nucleic acid sequences which encode the amino acid sequences
XX given in ABP35124 to ABP37960. The S. epidermidis sequences have
XX antibacterial activity and can be used in gene therapy. The sequences
XX can also be used in the diagnosis and treatment of bacterial infections,
XX particularly S. epidermidis infections. The sequences can be used to
XX screen for compounds able to interfere with the S. epidermidis life
XX cycle or inhibit S. epidermidis infection.
XX
XX CC N.B. The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from the
XX USPTO web site.
XX
XX SQ Sequence 438 AA:
Query Match 4.2%; Score 85.5; DB 23; Length 438;
Best Local Similarity 19.2%; Pred. No. 5.1;
Matches 62; Conservative 49; Mismatches 127; Indels 85; Gaps 13;
QY 32 ECGPTSTITTFNTRNAFEG--HYVYKGLXDQGRNDEGGRQVAGSLPFDSCNVARTRS 89
DB 154 ENGFTKIDPPIITLASAPESTSEIFHTKYFDEDAFLSQS-----QLYMEAAAHNGR- 205
QY 90 LMRGIFVTTVVYISHPLEVTKYDRAVR--YOCFYMEADKTVSAQIEVSEITTAFOQI 147
DB 206 -----VFSGPFRAEKSKTRRHLLIEFWMIEPMAFTNNAESLEQEOYVSHI 253
QY 148 VPMV--CRYEILDGPTGQPVQFALIGQPVYHKWTCDSVTDTFCVAVHSCFVDGNGD 205
DB 254 VOSVLNHCQLEK-----ALDRDT--TKLEKVAATPFPRISTD 289
QY 206 TVEILNADGKALDKYILNNLEFPTDLMAQEAHYKYADRSQLE--YOCQISITTEKPRS 263
DB 290 AIEFLKREG-----FDIEWGEDFGAPHEIALNHHYDLVPFTNNYPTKIKPFYMQPMP 342
QY 264 E-----CYRPCSEPCGCAVKTGGAAPAAAOQLRLKRSAPENIIDVRTDINTL 317
DB 343 DNEEDTVLCA--DLIAPEGGELIGGSEKIN-----DLELLEOR-----INEH 382
QY 318 EISDDNQLPVDLRRHALLQHNQ 340
```


GenCore version 5.1.4.P5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 24, 2003, 08:51:14 ; Search time 18 Seconds
(without alignments)
632.593 Million cell updates/sec

Title: US-10-054-562a-4

Perfect score: 2026

Sequence: 1 MMRLIAFCTTLIALSYIP.....ALIAAVITITSPKRNQKA 387

Scoring table: BLOSUM62

Searched: Gapop 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Issued Patents-AA:*
1: /cgn2_6/prodata/2/iaa/5A.COMB.pep:*
2: /cgn2_6/prodata/2/iaa/5B.COMB.pep:*
3: /cgn2_6/prodata/2/iaa/5A.COMB.pep:*
4: /cgn2_6/prodata/2/iaa/5B.COMB.pep:*
5: /cgn2_6/prodata/2/iaa/PCTUS.COMB.pep:*
6: /cgn2_6/prodata/2/iaa/backfilest1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2026	100.0	387	4	US-09-323-427-4
2	2026	100.0	387	4	US-09-812-642-4
3	1086.5	53.6	245	4	US-09-323-427-17
4	1086.5	53.6	245	4	US-09-812-642-17
5	776.5	38.3	271	4	US-09-323-427-9
6	776.5	38.3	271	4	US-09-812-642-9
7	100	4.9	935	2	US-07-707-367-2
8	96.5	4.8	989	2	US-08-070-301-16
9	87	4.3	3038	1	US-08-450-332-2
10	87	4.3	3038	2	US-08-637-640-2
11	87	4.3	3038	4	US-09-004-406C-2
12	85.5	4.2	438	4	US-09-134-001C-3139
13	83	4.1	1222	2	US-08-682-517-15
14	83	4.1	1252	2	US-08-682-517-9
15	81.5	4.0	1911	1	US-08-348-0068-5
16	81.5	4.0	1911	2	US-08-800-825A-5
17	81.5	4.0	1911	4	US-09-158-657-5
18	81.5	4.0	1911	5	PCT-US94-10166-5
19	80.5	4.0	430	1	US-08-785-076-3
20	80.5	4.0	430	4	US-09-018-824-3
21	80.5	4.0	432	1	US-08-785-076-2
22	80.5	4.0	432	4	US-09-018-824-2
23	78.5	3.9	434	2	US-08-484-575A-20
24	78.5	3.9	434	3	US-08-477-459-20
25	78.5	3.9	434	3	US-08-479-869-20
26	78.5	3.9	434	4	US-08-486-414-46
27	78.5	3.9	434	5	PCT-US94-01826A-20

28	78.5	3.9	434	5	PCT-US94-02252A-20	Sequence 20, Appl
29	78.5	3.9	434	5	PCT-US96-03916-11	Sequence 11, Appl
30	78.5	3.9	434	5	PCT-US96-03916-67	Sequence 67, Appl
31	78.5	3.9	570	2	US-08-484-993B-16	Sequence 16, Appl
32	78.5	3.9	570	2	US-08-484-158B-16	Sequence 16, Appl
33	78.5	3.9	570	2	US-08-484-596A-16	Sequence 16, Appl
34	78.5	3.9	570	2	US-08-480-150A-16	Sequence 16, Appl
35	78.5	3.9	570	3	US-08-458-731-16	Sequence 16, Appl
36	78.5	3.9	570	3	US-08-149-223A-16	Sequence 16, Appl
37	78	3.8	549	1	US-08-325-071-61	Sequence 61, Appl
38	78	3.8	549	1	US-08-461-004A-61	Sequence 61, Appl
39	78	3.8	650	1	US-08-325-071-63	Sequence 63, Appl
40	78	3.8	650	4	US-08-461-004A-63	Sequence 10, Appl
41	77.5	3.8	646	4	US-09-336-643A-10	Sequence 67, Appl
42	77.5	3.8	650	4	US-08-325-071-67	Sequence 67, Appl
43	77.5	3.8	650	4	US-08-461-004A-67	Sequence 27, Appl
44	77	3.8	4536	4	US-09-180-422B-27	Sequence 1, Appl
45	76.5	3.8	1184	2	US-08-918-914-1	

ALIGNMENTS

```
RESULT 1
US-09-323-427-4
; Sequence 4, Application US/09323427
; Patent No. 6248329
; GENERAL INFORMATION:
; APPLICANT: Chandrasekar, Ramaswamy
; APPLICANT: Morales, Tony H.
; TITLE OF INVENTION: Parasitic Helminth Cuticlin Proteins, Nucleic Acid
; TITLE OF INVENTION: Molecules, and Uses Thereof
; FILE REFERENCE: HN-8
; CURRENT APPLICATION NUMBER: US/09/323.427
; CURRENT FILING DATE: 1999-06-01
; EARLIER APPLICATION NUMBER: 60/087,435
; EARLIER FILING DATE: 1998-06-01
; NUMBER OF SEQ. ID NOS: 18
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 4
; LENGTH: 387
; TYPE: PRT
; ORGANISM: Dirofilaria immitis
US-09-323-427-4

Query Match      100.0%; Score 2026; DB 4; Length 387;
Best Local Similarity 100.0%; Pred. No. 4.3e-231;
Matches 387; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MMRLIAFCTTLIALSYIPVNDNGVEGEPEIECGPTSTTTFNTRNAFEGHVVYKGLYDQ 60
Db 1 MMRLIAFCTTLIALSYIPVNDNGVEGEPEIECGPTSTTTFNTRNAFEGHVVYKGLYDQ 60
QY 61 EGCNDGSGROVAGISLPFDSGNVARTSLNPRGIEFTTVVISFHPLEFKVDRAVRVQ 120
Db 61 EGCNDGSGROVAGISLPFDSGNVARTSLNPRGIEFTTVVISFHPLEFKVDRAVRVQ 120
QY 121 CFYEADKTSYSAQLEVESEITTAFTQIVPMPCVRCYETIELDGGPTGQPVQAFATIGOPVYHKW 180
Db 121 CFYEADKTSYSAQLEVESEITTAFTQIVPMPCVRCYETIELDGGPTGQPVQAFATIGOPVYHKW 180
QY 181 TCDSYVDFTCVAVHSCFVDDGNGDVEIILNADSCALDKYLLNNLEYPTDLMAGOEAAHVY 240
Db 181 TCDSYVDFTCVAVHSCFVDDGNGDVEIILNADSCALDKYLLNNLEYPTDLMAGOEAAHVY 240
QY 241 KYADRSOLFQOCQISTITKEPNSCVPRQCSPEGFAVVTGGAAPAAAOURLKKR 300
Db 241 KYADRSOLFQOCQISTITKEPNSCVPRQCSPEGFAVVTGGAAPAAAOURLKKR 300
QY 301 SAEPENTIDVTRDTINTLEISDDNQALPVDLRHRALLQHNQOPVILAVONGICMSPGFS 360
Db 301 SAEPENTIDVTRDTINTLEISDDNQALPVDLRHRALLQHNQOPVILAVONGICMSPGFS 360
```

Qy 361 MEMGSIALLAAVITTSKFRPNOKA 387
|
Db 361 MEMGSIALLAAVITTSKFRPNOKA 387

RESULT 2

US-09-812-642-4
; Sequence 4, Application US/09812642
; Patent No. 6368600
; GENERAL INFORMATION:
; APPLICANT: Chandrashekar, Ramaswamy
; APPLICANT: Morales, Tony H.
; TITLE OF INVENTION: Parasitic Helminth Cuticlin Proteins, Nucleic Acid
; FILE REFERENCE: HM-8
; CURRENT APPLICATION NUMBER: US/09/812,642
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: 09/323,427
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 387
; TYPE: PRT
; ORGANISM: Dirofilaria immitis
US-09-812-642-4

Query Match 100.0%; Score 2026; DB 4; Length 387;
Best Local Similarity 100.0%; Pred. No. 4,3e-231;
Matches 387; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MMIRLAFCTTLIALSY-SPVDNGVEGEPEIECGPTSTINFTNRNAPFGHYVKGLYDQ 60
|
Db 1 MMIRLAFCTTLIALSY-SPVDNGVEGEPEIECGPTSTINFTNRNAPFGHYVKGLYDQ 60
|
Qy 61 ECCRDEGGROYAGISLPPDSCNVAARTSLNPRGIFVTTTVVISFHPLEVTVDRAVRYO 120
|
Db 61 ECCRDEGGROYAGISLPPDSCNVAARTSLNPRGIFVTTTVVISFHPLEVTVDRAVRYO 120
|
Qy 121 CFYMEADKTVSAQIEVSEITTAFOQIYVMPVCRYEILDGGPTGQVQVPAIIGOPVYHNM 180
|
Db 121 CFYMEADKTVSAQIEVSEITTAFOQIYVMPVCRYEILDGGPTGQVQVPAIIGOPVYHNM 180
|
Qy 181 TCDSETVDFCAVHSCFVDGNGDTVELLNADGALDKYLLNNLEYPTDLMAGOEAHNY 240
|
Db 181 TCDSETVDFCAVHSCFVDGNGDTVELLNADGALDKYLLNNLEYPTDLMAGOEAHNY 240
|
Qy 241 KYADESOLFYOQOISITTIKEPNSCEVRCPOCSPOGFGAAYKTGGAAPAAQOLRLKKR 300
|
Db 241 KYADESOLFYOQOISITTIKEPNSCEVRCPOCSPOGFGAAYKTGGAAPAAQOLRLKKR 300
|
Qy 301 SAEPENIIDVTDINTLEISDNOALPYDLRHRALLQHNQGPVILAAVONGICMSPFGFS 360
|
Db 301 SAEPENIIDVTDINTLEISDNOALPYDLRHRALLQHNQGPVILAAVONGICMSPFGFS 360
|
Qy 361 MEMGSIALLAAVITTSKFRPNOKA 387
|
Db 361 MEMGSIALLAAVITTSKFRPNOKA 387

RESULT 3

US-09-323-427-17
; Sequence 17, Application US/09323427
; Patent No. 6248329
; GENERAL INFORMATION:
; APPLICANT: Chandrashekar, Ramaswamy
; APPLICANT: Morales, Tony H.
; TITLE OF INVENTION: Parasitic Helminth Cuticlin Proteins, Nucleic Acid
; FILE REFERENCE: HM-8
; CURRENT APPLICATION NUMBER: US/09/323,427
; CURRENT FILING DATE: 1999-06-01
; EARLIER APPLICATION NUMBER: 60/087,435

; EARLIER FILING DATE: 1998-06-01
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 245
; TYPE: PRT
; ORGANISM: Brugia malayi
US-09-323-427-17

Query Match 53.6%; Score 1086.5; DB 4; Length 245;
Best Local Similarity 80.6%; Pred. No. 3,8e-120;
Matches 195; Conservative 24; Mismatches 22; Indels 1; Gaps 1;

Qy 3 IRLIAFCTTLIALSY-SIPVDNGVEGEPEIECGPTSTINFTNRNAPFGHYVKGLYDQ 61
|
Db 4 MOICFSLSYMIILASINAIPIIDNGVESEPEIECGPTSTINFTNRNAPFGHYVAKLYGNSQ 63
|
Qy 62 GCRNDEGGROYAGISLPPDSCNVAARTSLNPRGIFVTTTVVISFHPLEVTVDRAVRYOC 121
|
Db 62 GCRNDEGGROYAGISLPPDSCNVAARTSLNPRGIFVTTTVVISFHPLEVTVDRAVRYOC 121
|
Qy 122 FYMEADKTVSAQIEVSEITTAFOQIYVMPVCRYEILDGGPTGQVQVPAIIGOPVYHNM 181
|
Db 122 FYMEADKTVSAQIEVSEITTAFOQIYVMPVCRYEILDGGPTGQVQVPAIIGOPVYHNM 181
|
Qy 182 CDSEIVDFCAVHSCFVDGNGDTVELLNADGALDKYLLNNLEYPTDLMAGOEAHNYK 241
|
Db 182 CDSEIVDFCAVHSCFVDGNGDTVELLNADGALDKYLLNNLEYPTDLMAGOEAHNYK 241
|
Qy 242 YA 243
|
Db 244 YA 245

RESULT 4

US-09-812-642-17
; Sequence 17, Application US/09812642
; Patent No. 6368600
; GENERAL INFORMATION:
; APPLICANT: Chandrashekar, Ramaswamy
; APPLICANT: Morales, Tony H.
; TITLE OF INVENTION: Parasitic Helminth Cuticlin Proteins, Nucleic Acid
; FILE REFERENCE: HM-8
; CURRENT APPLICATION NUMBER: US/09/812,642
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: 09/323,427
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 245
; TYPE: PRT
; ORGANISM: Brugia malayi
US-09-812-642-17

Query Match 53.6%; Score 1086.5; DB 4; Length 245;
Best Local Similarity 80.6%; Pred. No. 3,8e-120;
Matches 195; Conservative 24; Mismatches 22; Indels 1; Gaps 1;

Qy 3 IRLIAFCTTLIALSY-SIPVDNGVEGEPEIECGPTSTINFTNRNAPFGHYVKGLYDQ 61
|
Db 4 MOICFSLSYMIILASINAIPIIDNGVESEPEIECGPTSTINFTNRNAPFGHYVAKLYGNSQ 63
|
Qy 62 GCRNDEGGROYAGISLPPDSCNVAARTSLNPRGIFVTTTVVISFHPLEVTVDRAVRYOC 121
|
Db 62 GCRNDEGGROYAGISLPPDSCNVAARTSLNPRGIFVTTTVVISFHPLEVTVDRAVRYOC 121
|
Qy 122 FYMEADKTVSAQIEVSEITTAFOQIYVMPVCRYEILDGGPTGQVQVPAIIGOPVYHNM 181
|
Db 122 FYMEADKTVSAQIEVSEITTAFOQIYVMPVCRYEILDGGPTGQVQVPAIIGOPVYHNM 181
|
Qy 182 CDSEIVDFCAVHSCFVDGNGDTVELLNADGALDKYLLNNLEYPTDLMAGOEAHNYK 241
|
Db 182 CDSEIVDFCAVHSCFVDGNGDTVELLNADGALDKYLLNNLEYPTDLMAGOEAHNYK 241

QY 328 VDLRH-RALLQHGQPVIL-----AAVONGICMSPF 357
 DB 1529 KTMARHSILKPGGMVILITHEKHTRLGIFGLFADWAGVDDGRCTEPF 1580

RESULT 11

US-09-004-406C-2
 ; Sequence 2, Application US/09004406C
 ; Patent No. 6174706
 ; GENERAL INFORMATION:
 ; APPLICANT: Vincel, Victor A.
 ; APPLICANT: Conder, Michael J.
 ; APPLICANT: McAda, Phyllis C.
 ; APPLICANT: Reeves, Christopher D.
 ; APPLICANT: Rambosek, John
 ; APPLICANT: Davis, Charles Ray
 ; APPLICANT: Hendrickson, Lee E.
 ; TITLE OF INVENTION: DNA ENCODING TRIOL POLYKETIDE SYNTHASE
 ; FILE REFERENCE: 19076PDA
 ; CURRENT APPLICATION NUMBER: US/09/004,406C
 ; CURRENT FILING DATE: 1998-01-08
 ; PRIOR APPLICATION NUMBER: 08/637,640
 ; PRIOR FILING DATE: 1996-08-23
 ; PRIOR APPLICATION NUMBER: 08/148,132
 ; PRIOR FILING DATE: 1993-11-02
 ; NUMBER OF SEQ ID NOS: 41
 ; SOFTWARE: FASTSEQ for Windows Version 4.0
 ; SEQ ID NO 2
 ; LENGTH: 3038
 ; TYPE: PRT
 ; ORGANISM: TPKS Protein
 ; US-09-004-406C-2

Query Match 4.3%; Score 87; DB 4; Length 3038;
 Best Local Similarity 19.7%; Pred. No. 6.7; Mismatches 138; Indels 186; Gaps 24;
 Matches 93; Conservative 55;

QY 6 IAFCTTLALSYSPVNDGVEGEPEIEC--GPTSI-----TINENT 44
 DB 1175 VAFQVIGV--YSSP-----GDRRLRCLYPTHVDRITLYPSLCLATAESGCEVAFYT 1226
 QY 45 RAAFGHYVVGK---LYDEGCRNDEGGNOVAGISL--PPDSCNVAKTSLNRRGIF--VT 98
 DB 1227 INTYKRGYLSGDIVVFAE---QTLFQYENITFKPSPDASTDHAFRMSWGPILT 1282
 QY 99 TTVVLSFPHLEFVTKVDR-----AYRQCFYMEADKTVSAQIEVSEITTA---FOQIYPM 150
 DB 1283 PDSLNDNEYATADQDKKAIPIETRIYFYI---RSFLSQLLEERQQAHLQKQIEWL 1339
 QY 151 PVCRYEILDGPTGPVQFAIGQPVYHKWTCSETVDITFCVAVHSCFVDD-----201
 DB 1340 E-----QVLASAKGR-----HLWYDPGWNEDTEAQIEHLCTANSYHPRVLQ 1383
 QY 202 -----GNGDTVELNADGALDKYILNN-----LEYPPIDLAGQPAHYKYAD 244
 DB 1394 RVGQHLLEPTVRSGNPNPDLDDHG-LTLEFYNTLSFGPALHYAREIVA-QIAHRYQSM 1441
 QY 245 RSLQFYQCOISITIKPEPSECVRPQCSPEQFGAVKGTGAAPAAQOL-----294
 DB 1442 ILLEI-----GAG---TGATKYVLAATPOLGNSYTYDI 1472
 QY 295 -----RULKRSAPENIIDVRTDITLLEISDNOALP 327
 DB 1473 STGFPEQAREQFAPPEDMVFEPLDIRSPAEQGEF-HAYDLIIASNVLAHTPD---LE 1528
 QY 328 VDLRH-RALLQHGQPVIL-----AAVONGICMSPF 357
 DB 1529 KTMARHSILKPGGMVILITHEKHTRLGIFGLFADWAGVDDGRCTEPF 1580

RESULT 12
 US-09-134-001C-3139

; Sequence 3139, Application US/09134001C
 ; Patent No. 6380370
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn Doucette-Stamm et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
 ; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: GTC-007
 ; CURRENT APPLICATION NUMBER: US/09/134,001C
 ; CURRENT FILING DATE: 1998-08-13
 ; PRIOR APPLICATION NUMBER: US 60/064,964
 ; PRIOR FILING DATE: 1997-11-08
 ; PRIOR APPLICATION NUMBER: US 60/055,779
 ; PRIOR FILING DATE: 1997-08-14
 ; NUMBER OF SEQ ID NOS: 5674
 ; SEQ ID NO 3139
 ; LENGTH: 438
 ; TYPE: PRT
 ; ORGANISM: Staphylococcus epidermidis
 ; US-09-134-001C-3139

Query Match 4.2%; Score 85.5; DB 4; Length 438;
 Best Local Similarity 19.2%; Pred. No. 0.37;
 Matches 62; Conservative 49; Mismatches 127; Indels 85; Gaps 13;

QY 32 ECGPSITINENTRAFEG--HYVYKGLYDQSGCRNDEGGNOVAGISLPPDSCNVAKTS 89
 DB 154 ENGFTKIDPRLTASAPESGTSLEFHTKYFDEDAFLSQS-----QLYMEAAAAMHGR- 205
 QY 90 LNPGRGIVTTTVISFHLFVTKVDRAVR--VQCFYMEADKTVSAQIEVSEITTAFOQI 147
 DB 206 -----VFSGFPRFAEKSTRRHLLEFWNIEPMFTNHAESLEIOEQYVSHI 253
 QY 148 VPMRV--CRYEILDGPTGPVQFAIGQPVYHKWTCSETVDITFCVAVHSCFVDDGND 205
 DB 254 VQSVLNHQLLEK-----ALBQDT--TKLEKATFPRIAYSD 289
 QY 206 TVEILNADGALDKYILNNLEYPTDLMAQGAHYKYADRSOLF--YQCOISITIKPEPS 263
 DB 290 ALEFLKREG-----FDIEWGEDFGAPHTAIANNHYDLPVFTINYPRKIKFYWQPNP 342
 QY 264 E-----CYRQCSPEQFGAVKGTGAAPAAQOLRLKRSAPENIIDVRTDINTL 317
 DB 343 DNEEDTVLCA--DLIAPEGGELIGSERIN-----DLELLEQR-----INER 382
 QY 318 EISDNOALPVDLRHALLQHG 340
 DB 383 ELDEESYIYDLRRKGVSPHSG 405

RESULT 13

US-08-682-517-15
 ; Sequence 15, Application US/08682517
 ; Patent No. 5874267
 ; GENERAL INFORMATION:
 ; APPLICANT:
 ; TITLE OF INVENTION: Expression of surface layer proteins
 ; NUMBER OF SEQUENCES: 25
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/682,517
 ; FILING DATE:

CLASSIFICATION:
 ; INFORMATION FOR SEQ ID NO: 15:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1222 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-682-517-15

Query Match	4.03;	Score 81.5;	DB 4;	Length 1911;
Best Local Similarity	21.58;	Prod. No. 14;		
Matches 84;	Conservative 47;	Mismatches 137;	Indels 123;	Gaps
QY 27	GEPEI-ECGPSIVINENTRNAEFGHYV-----KGLY----	DOEGCRNDEGGRQ--	71	
Db 324	GTPVTEWTAISITITWDSGPNPDVSYVYIEYKSKSDGPGYQIKEDITTTIRYSIGGLSPN	383		
QY 72	-----VAGISL-----PPDSCNART-----RSLNPGRIPTVTTVVISFHPLEVTK	112		
	: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :			
Db 384	SEVEIWSAVNSIGCGPPSESVVVRTGEQAPARPBRVQARMISATMTIIVQEEPEPBG	443		
QY 113	VDRAYVOCFYMADKTV-----SAQIEVSETTA---FOQIYMPRCREIILDDGPTQG	165		
	: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :			
Db 444	LIRGTRYV-YIMEEHHYGNQKKNVDDSLITITGSLLEDETYTVKRLATSVGGDPLSD	502		
QY 166	PVQF---AIIQGPVYHKWTCDSFTVDFC-----AVV--HSCFVDDNGDGTVELTA	212		
	: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :			
Db 503	PIQVKTQGVGCGPMNLRARASGETSITLWSPPROESITIKYELLFREGDHGREV-----	557		
QY 213	DGCALD---KLLNLEVPDTLMAGQEAHYKYVDRSQRLEFGQOISITTIKPRSECVRPQ	269		
Db 558	-GRFFDPTTSIVYEDLKPNE-----YAFRLAARS-----	566		
QY 270	CSEPGFGAVYTGGAAPAAQAOLRLIKKRSAPENI---IDVKTIDINTLEISDDNOLP	327		
Db 587	---PQGLGAF-----TPVVRQRTIQSKRSAPQGVKCVSR-----STALL	624		
QY 328	VDLRHRALLQHNQGPVILIAAVQNCIGCSPPG	358		
Db 625	VSMRPPPETING-----ALGVSVRYRPLG	650		

```

1      RESULT 18
2      PCT-US94-10166-5
3      Sequence 5, Application PC/TUS9410166
4      GENERAL INFORMATION:
5      APPLICANT: RODAN, GIDEON A
6      APPLICANT: SCHMIDT, AZRIEL
7      APPLICANT: RUTLEDGE, SU JANE
8      TITLE OF INVENTION: CDNA ENCODING A NOVEL HUMAN PROTEIN
9      TITLE OF INVENTION: TYROSINE PHOSPHATASE
10     NUMBER OF SEQUENCES: 6
11     CORRESPONDENCE ADDRESSES:
12     ADDRESSEE: JOHN W. WALLEN III
13     STREET: P O. BOX 2000, 126 E. LINCOLN AVE.
14     CITY: RAHWAY
15     STATE: NJ
16     COUNTRY: USA
17     ZIP: 07065
18     COMPUTER READABLE FORM:
19     MEDIUM TYPE: Floppy disk
20     COMPUTER: IBM PC compatible
21     OPERATING SYSTEM: PC-DOS/MS-DOS
22     SOFTWARE: PatentIn Release #1.0, Version #1.25
23     CURRENT APPLICATION DATA:
24     APPLICATION NUMBER: PCT/US94/10166
25     FILING DATE: 09-SEPT-1994
26     CLASSIFICATION:
27     PRIOR APPLICATION DATA:
28     APPLICATION NUMBER: US 08/122,032
29     FILING DATE: 14-SEP-1993
30     ATTORNEY/AGENT INFORMATION:
31     NAME: WALLEN, JOHN W III
32     REGISTRATION NUMBER: 35403
33     REFERENCE/DOCKET NUMBER: 18992
34     TELECOMMUNICATION INFORMATION:
35     TELEPHONE: 908-594-3905
36     TELEFAX: 908-594-4720
37     TELEX: 138825
38     INFORMATION FOR SEQ ID NO: 5:
39     SEQUENCE CHARACTERISTICS:

```

```

: LENGTH: 1911 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: PCT-US94-10166-5

Query Match      4.0%; Score 81.5; DB 5; Length 1911;
Best Local Similarity 21.5%; Pred. No. 14;
Matches 84; Conservative 47; Mismatches 137; Indels 123; Gaps 20;

OY 27 GEPET-ECGPGSITINFTNRNAAFEHVVV-----RGLY---DQSGCRNDEGGRQ-- 71
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 324 GTPWTEMTATISITITWDSGNDPDSYVIEYKSKSGQGPQIQKEDITTTKSTIGLSFN 383
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 72 -----VAGISL---PDSQCNVART-----RSLNPGIFVTTTVVISFHPLFYTK 112
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 384 SEYEIWSAVNSISIGGPSESVYTRTGBQAPARPRNVQARMLSATMTIVQEEPEVBNQ 443
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 113 VDRAYRQCCFVMEADKY-----SAQIEVSEITTA---FQTQIVPMPYCRYEILDGSPGQ 165
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 444 LIRGRYV-YTMEPRHPAGNMOKHNVDSLLTTVGSLLEDEYTYTRVLAFTSVSGDPLSD 502
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 166 PVQF-----AIIIGOPVYHKWTCDSFETVDFQC-----AVV--HSGCFVDGNGNDFVEILNA 212
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 503 PIQVKTQCGVCGQPNLPAEARSETITLMSPPQESTIKYELLFREGDHGREY----- 557
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 213 DGCALD---KYLNLNLEYPTDLMAQEAHVYKYADRSQLEFYQOQISITIKENSECVRPQ 269
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 558 -GRTEPDPTSTVYEDLRKNTD-----YAFRLARS----- 586
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 270 CSEPGFCFAYVTGCAAKPAAQAOLRLKKSAPENI--IDVRPDINTLEISDNOALP 327
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 587 ---PGLGAF-----TPVVRQRTLOSKSPAPQDQVCVSVR-----STAIL 624
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 328 VDLRRHALLHNGQPVILAAVONGICMSPFQ 358
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 625 VSMRPPPPETHNG-----ALVGYSVRYRPLG 650
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 19
US-08-785-076-3
: Sequence 3, Application US/08785076
: Patent No. 5789217
: GENERAL INFORMATION:
: APPLICANT: Hodgson, John
: APPLICANT: Lawlor, Elizabeth
: TITLE OF INVENTION: No. 5789217el tRNA Synthetase
: NUMBER OF SEQUENCES: 3
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: SmithKline Beecham Corporation
: STREET: 709 Swedeland Road
: CITY: King of Prussia
: STATE: PA
: COUNTRY: USA
: ZIP: 19406-0939
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/785,076
: FILING DATE: 17-JAN-1997
: CLASSIFICATION: 514
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 9601096.2
: FILING DATE: 19-JAN-1996
: APPLICATION NUMBER: 9615845.6
: FILING DATE: 27-JUL-1996
: APPLICATION NUMBER: 9622617.0
: FILING DATE: 30-OCT-1996
: ATTORNEY/AGENT INFORMATION:

```

```
NAME: Glimmi, Edward R
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P31354-3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ. ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 430 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-785-076-3

Query Match
Best local Similarity 19.4%; Score 80.5; DB 1; Length 430;
Matches 55; Conservative 46; Mismatches 103; Indels 79; Gaps 14;

QY 102 VISFHPLEVTYKVDRAVR--VQCFYMEADKTVSAQIEVSEITTAFOQTQIYP--MPVCRYEI 157
| | | | | : : : : : | : : : : | : : : : | : : : : | : : : : |
DB 198 VFSEGFPTFRAEKSKTRHLLIEFMWIEGEMAFTHNAESLEIOEQYVTHVVKSVLENCKLEL 257
| | | | | : : : : : | : : : : | : : : : | : : : : | : : : : |
QY 158 LDGSGPTGPQVFAIIIGQPYVHKWTCSEYVDY-FCANVHSCFVDDGNGDTVEILNADGCA 216
| | | | | : : : : : | : : : : | : : : : | : : : : | : : : : |
DB 258 -----KILERDTSKLEKVAATPPRISYD-----DAIEFLKAE-- 290
| | | | | : : : : : | : : : : | : : : : | : : : : | : : : : |
QY 217 LDKYLLNLEYPPTDLMAGOEAHVYKYADRSOLF--YQCOISITIKPNSECVAP--OCSE- 272
| | | | | : : : : : | : : : : | : : : : | : : : : | : : : : |
DB 291 ----FDIEWGEDFGAPHETAIANHYDLPVFTTNYPTKIKRPYMQPNENEETVLCADL 345
| | | | | : : : : : | : : : : | : : : : | : : : : | : : : : |
QY 273 --PQSGAVKVTGGAAKPAAAOQLRLKKRSAPENIIDVRDINPLESDNOALPVLD 330
| | | | | : : : : : | : : : : | : : : : | : : : : | : : : : |
DB 346 IAEYGGEI-IGGS---ERVDLLELEQRYVEH-----GLDEEAYSYYLDL 387
| | | | | : : : : : | : : : : | : : : : | : : : : | : : : : |
QY 331 RHRALLQHGQPYIIAAYONGICMSPFGFSPMGSLIALIAAY 373
| | | | | : : : : : | : : : : | : : : : | : : : : | : : : : |
DB 388 RRYGSVPH-----C-----GFGLGERTVAMISGV 412
| | | | | : : : : : | : : : : | : : : : | : : : : | : : : : |

RESULT 20
US-09-018-824-3
; Sequence 3, Application US/09018824
; Patent No. 6410286
; GENERAL INFORMATION:
; APPLICANT: Hodgson, John
; APPLICANT: Lawlor, Elizabeth
; TITLE OF INVENTION: No. 6410286el tRNA Synthetase
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/018.824
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER: 08/785.076
; FILING DATE: 17-JAN-1997
; APPLICATION NUMBER: 9601096.2
; FILING DATE: 19-JAN-1996
; APPLICATION NUMBER: 9615845.6
; FILING DATE: 27-JUL-1996
```

```
APPLICATION NUMBER: 9622617.0
FILING DATE: 30-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Glimmi, Edward R
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P31354-3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ. ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 430 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-018-824-3

Query Match
Best local Similarity 19.4%; Score 80.5; DB 4; Length 430;
Matches 55; Conservative 46; Mismatches 103; Indels 79; Gaps 14;

QY 102 VISFHPLEVTYKVDRAVR--VQCFYMEADKTVSAQIEVSEITTAFOQTQIYP--MPVCRYEI 157
| | | | | : : : : : | : : : : | : : : : | : : : : | : : : : |
DB 198 VFSEGFPTFRAEKSKTRHLLIEFMWIEGEMAFTHNAESLEIOEQYVTHVVKSVLENCKLEL 257
| | | | | : : : : : | : : : : | : : : : | : : : : | : : : : |
QY 158 LDGSGPTGPQVFAIIIGQPYVHKWTCSEYVDY-FCANVHSCFVDDGNGDTVEILNADGCA 216
| | | | | : : : : : | : : : : | : : : : | : : : : | : : : : |
DB 258 -----KILERDTSKLEKVAATPPRISYD-----DAIEFLKAE-- 290
| | | | | : : : : : | : : : : | : : : : | : : : : | : : : : |
QY 217 LDKYLLNLEYPPTDLMAGOEAHVYKYADRSOLF--YQCOISITIKPNSECVAP--OCSE- 272
| | | | | : : : : : | : : : : | : : : : | : : : : | : : : : |
DB 291 ----FDIEWGEDFGAPHETAIANHYDLPVFTTNYPTKIKRPYMQPNENEETVLCADL 345
| | | | | : : : : : | : : : : | : : : : | : : : : | : : : : |
QY 273 --PQSGAVKVTGGAAKPAAAOQLRLKKRSAPENIIDVRDINPLESDNOALPVLD 330
| | | | | : : : : : | : : : : | : : : : | : : : : | : : : : |
DB 346 IAEYGGEI-IGGS---ERVDLLELEQRYVEH-----GLDEEAYSYYLDL 387
| | | | | : : : : : | : : : : | : : : : | : : : : | : : : : |
QY 331 RHRALLQHGQPYIIAAYONGICMSPFGFSPMGSLIALIAAY 373
| | | | | : : : : : | : : : : | : : : : | : : : : | : : : : |
DB 388 RRYGSVPH-----C-----GFGLGERTVAMISGV 412
| | | | | : : : : : | : : : : | : : : : | : : : : | : : : : |

RESULT 21
US-08-785-076-2
; Sequence 2, Application US/08785076
; Patent No. 5789217
; GENERAL INFORMATION:
; APPLICANT: Hodgson, John
; APPLICANT: Lawlor, Elizabeth
; TITLE OF INVENTION: No. 5789217el tRNA Synthetase
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/785.076
; FILING DATE: 17-JAN-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION NUMBER: 9601096.2
; FILING DATE: 19-JAN-1996
; APPLICATION NUMBER: 9615845.6
```


CURRENT APPLICATION DATA:


```

; REFERENCE/DOCKET NUMBER: 39116-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 434 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US96-03916-11

Query Match
Best Local Similarity 18.9%; Pred. No. 2.5; DB 5; Length 434;
Matches 80; Conservative 55; Mismatches 148; Indels 141; Gaps 22;

QY 45 RNAFECHVYVYKGLYDQECRNDEGROVAGISLPFDSQNVARTSLNPRGIFVTTTVIS 104
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 55 RNAMDRHLFLRNAFWITVLSSFASQSTAATYDY-----ILGRALDALTIPIAVG 105

QY 105 FHPLFVTKVDRAVYVOCFYME-----ADKTVSAQIEVSEITTAFTQTQIYMPVCRYEIL 158
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 106 PYNRYLTRVSRG---CDVVELNPISNVDDMISAKEKER-----141

QY 159 DGGP-TGQPVQFATII---GQ---PYVHK--WTCDSFTVDTFCAY-----VHSC 197
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 142 -GGPFESVWFYVYIKGDDDEDKCPIYRKREYRCGVDQLSECAVSAQMAVADYVPST 200

QY 198 FVDDNGDVEILLNADGALDKYLL-----222
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 201 LV-SRNAGGLTFSPPTAALSGQYLLTLKIGFAQTALVLEVNDRLCKISQLNPLPSKC 259

QY 223 -NNLEPTDLMAQGEAHVYKADRS-----QLEYQOCISITIKEPNSECVRPQ 269
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 260 WTEQYOTGF--QGEHLVYPIADNTNRHADVDYRGYEDILQRMNNLLRKKNPSPADPRP 316

QY 270 CSEPOGGAVKYTGGAAPAAQRLRLKRSAPENIIDVRTDINTLEISDDMQA---- 325
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 317 -SVQEIPIAVYTKKAGRTPDAS-----SEKKAPPE--DSEDDMQA-EASGENPALPE 366

QY 326 ---LPVDLRHRLIQ---HNGQPVIIAAVQ---NGICSPFGFSFMGLSIALIAAV 373
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 367 DDEVPEDTEHDDPNSDPDYNDMPAVIPVETTKSSNAVSM-PI-FAAFVACAVALVGLL 424

QY 374 IITIT 377
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 425 VWSI 428

RESULT 30
PCT-US96-03916-67
; Sequence 67, Application PC/TUS9603916
; GENERAL INFORMATION:
; APPLICANT: Wild, Martha A.
; APPLICANT: Cochran, Mark D.
; TITLE OF INVENTION: RECOMBINANT INFECTIOUS LARYNGOTRACHEITIS VIRUS
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/03916
; FILING DATE: 23-MAR-1995

```

```

; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/126,597
; FILING DATE: 24-SEP-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 39116-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 67:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 434 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US96-03916-67

Query Match
Best Local Similarity 18.9%; Pred. No. 2.5; DB 5; Length 434;
Matches 80; Conservative 55; Mismatches 148; Indels 141; Gaps 22;

QY 45 RNAFECHVYVYKGLYDQECRNDEGROVAGISLPFDSQNVARTSLNPRGIFVTTTVIS 104
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 55 RNAMDRHLFLRNAFWITVLSSFASQSTAATYDY-----ILGRALDALTIPIAVG 105

QY 105 FHPLFVTKVDRAVYVOCFYME-----ADKTVSAQIEVSEITTAFTQTQIYMPVCRYEIL 158
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 106 PYNRYLTRVSRG---CDVVELNPISNVDDMISAKEKER-----141

QY 159 DGGP-TGQPVQFATII---GQ---PYVHK--WTCDSFTVDTFCAY-----VHSC 197
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 142 -GGPFESVWFYVYIKGDDDEDKCPIYRKREYRCGVDQLSECAVSAQMAVADYVPST 200

QY 198 FVDDNGDVEILLNADGALDKYLL-----222
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 201 LV-SRNAGGLTFSPPTAALSGQYLLTLKIGFAQTALVLEVNDRLCKISQLNPLPSKC 259

QY 223 -NNLEPTDLMAQGEAHVYKADRS-----QLEYQOCISITIKEPNSECVRPQ 269
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 260 WTEQYOTGF--QGEHLVYPIADNTNRHADVDYRGYEDILQRMNNLLRKKNPSPADPRP 316

QY 270 CSEPOGGAVKYTGGAAPAAQRLRLKRSAPENIIDVRTDINTLEISDDMQA---- 325
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 317 -SVQEIPIAVYTKKAGRTPDAS-----SEKKAPPE--DSEDDMQA-EASGENPALPE 366

QY 326 ---LPVDLRHRLIQ---HNGQPVIIAAVQ---NGICSPFGFSFMGLSIALIAAV 373
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 367 DDEVPEDTEHDDPNSDPDYNDMPAVIPVETTKSSNAVSM-PI-FAAFVACAVALVGLL 424

QY 374 IITIT 377
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 425 VWSI 428

RESULT 31
US-08-484-993B-16
; Sequence 16, Application US/08484993B
; Patent No. 5837497
; GENERAL INFORMATION:
; APPLICANT: Harris Ph.D., Jeffrey D.
; APPLICANT: Hsu, Kuang T.
; APPLICANT: Podolski, Joseph S.
; TITLE OF INVENTION: Materials and Methods for Immunoreception
; NUMBER OF SEQUENCES: 59
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402

```


TITLE OF INVENTION: Materials and Methods for Immunoprecipitation

NUMBER OF SEQUENCES: 59

```

CORRESPONDENCE ADDRESS:
ADDRESS: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/458,731
FILING DATE: 09-NOV-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/012,990
FILING DATE: 29-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/973,341
FILING DATE: 09-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36,107
REFERENCE/DOCKET NUMBER: 31745
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6653
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 570 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-458-731-16

Query Match
Best Local Similarity 21.1%; Score 78.5; DB 3; Length 570;
Matches 63; Conservative 38; Mismatches 109; Indels 89; Gaps 16;

QY 17 YSLPVDNGVEGEPEIECGTSTITNTNNAFEGHVVYKGLYDQEGCRNDEGROVAGIS 76
DB 230 FSIKAVSRNVTSP-----LLLN-SRLAIG-----KDRE-CNPKATRAFALEF 271
QY 77 LPFDSQNAVARTSLNPRGIFVTTVVISPHPLFVTKVDRA-----RVQCRY 123
DB 272 PFNNSCGTTR-----WYTGDAVYENELVAARDVFTWSHGSITRDSIFRLRVSCSY 322
QY 124 MEADTVSAQIEVSEITTA-FQTOIVPMV-----CRYEILDGPGTGPVQPAI 171
DB 323 SVRSNAFPLSVQVFTIPPHLKTQHGRLLELKIADKHYSYITGD-----YV-VKL 376
QY 172 IGPVYHKWTCSEVDFPCAVVHSCFVDDG-NGDTEV-----ILNADGALDKYLLNLEY 227
DB 377 LRDPIYEVSVIRHRTDPSLIGLLHNCMATPGKNSQSLSQWPL-VKGC---PYVGDN--Y 430
QY 228 PTDLMAQGEA-----HYVYADRSQLFYQCQISITIKENNSC 265
DB 431 QTOILPVQKALDTPPSTYKRFSTFTSFVDTMAKMLRGVYLHCNVISICOPAGTSSC 489

RESULT 36
US-08-149-223A-16
Sequence 16, Application US/08149223A
Patent No. 6027727
GENERAL INFORMATION:
APPLICANT: Harris Ph.D., Jeffrey D.
APPLICANT: Hsu, Kuang T.
APPLICANT: Podolski, Joseph S.
TITLE OF INVENTION: Materials and Methods for Immunocontraception
NUMBER OF SEQUENCES: 59

```

```

CORRESPONDENCE ADDRESS:
ADDRESS: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/149,223A
FILING DATE: 09-NOV-1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/012,990
FILING DATE: 29-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/973,341
FILING DATE: 09-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36,107
REFERENCE/DOCKET NUMBER: 31745
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6653
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 570 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-149-223A-16

Query Match
Best Local Similarity 21.1%; Score 78.5; DB 3; Length 570;
Matches 63; Conservative 38; Mismatches 109; Indels 89; Gaps 16;

QY 17 YSLPVDNGVEGEPEIECGTSTITNTNNAFEGHVVYKGLYDQEGCRNDEGROVAGIS 76
DB 230 FSIKAVSRNVTSP-----LLLN-SRLAIG-----KDRE-CNPKATRAFALEF 271
QY 77 LPFDSQNAVARTSLNPRGIFVTTVVISPHPLFVTKVDRA-----RVQCRY 123
DB 272 PFNNSCGTTR-----WYTGDAVYENELVAARDVFTWSHGSITRDSIFRLRVSCSY 322
QY 124 MEADTVSAQIEVSEITTA-FQTOIVPMV-----CRYEILDGPGTGPVQPAI 171
DB 323 SVRSNAFPLSVQVFTIPPHLKTQHGRLLELKIADKHYSYITGD-----YV-VKL 376
QY 172 IGPVYHKWTCSEVDFPCAVVHSCFVDDG-NGDTEV-----ILNADGALDKYLLNLEY 227
DB 377 LRDPIYEVSVIRHRTDPSLIGLLHNCMATPGKNSQSLSQWPL-VKGC---PYVGDN--Y 430
QY 228 PTDLMAQGEA-----HYVYADRSQLFYQCQISITIKENNSC 265
DB 431 QTOILPVQKALDTPPSTYKRFSTFTSFVDTMAKMLRGVYLHCNVISICOPAGTSSC 489

RESULT 37
US-08-325-071-61
Sequence 61, Application US/08325071
Patent No. 5587311
GENERAL INFORMATION:
APPLICANT: COBON, Stewart Gary
APPLICANT: MOORE, Joanna Terry
APPLICANT: JOHNSON, Law Anthony York
APPLICANT: WILLIAMS, Peter
APPLICANT: KEMP, David Harold

```


APPLICANT: SRISKANTHA, Alagacone
APPLICANT: RIDING, George Alfred
APPLICANT: RAND, Keith No. 558731man
TITLE OF INVENTION: DNA Encoding A Cell Membrane
TITLE OF INVENTION: Glycoprotein Of A Tick Gut
NUMBER OF SEQUENCES: 71
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W.
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/325,071
FILING DATE: 14-OCT-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/062,109
FILING DATE: 17-MAY-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/926,368
FILING DATE: 07-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/242,196
FILING DATE: 06-JUL-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU87/00401
FILING DATE: 27-NOV-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU P14912
FILING DATE: 16-OCT-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU P12570
FILING DATE: 19-JUN-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU P9196
FILING DATE: 27-NOV-1986
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
TELEPHONE: 202 672 5300
TELEFAX: 202 672 5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 61:
SEQUENCE CHARACTERISTICS:
LENGTH: 549 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-325-071-61

Query Match 3.88; Score 78; DB 1; Length 549;
Best Local Similarity 23.48; Pred. No. 4.2;
Matches 62; Conservative 26; Mismatches 95; Indels 82; Gaps 15;

QY 115 RAYRVOG---PYMDADTVSAQIEVSEITTAFOFQIV-PMVCHYE--IIDSGPTGPGV 167
Db 146 KAYECTCPRGTVADGTTCSISHTVSCIAEQKOTCAPTEDCKVHKGTVLCCEPMNO-- 203
QY 168 QFALIGQPVYHKWTCDSEFVDTFCFVAVHSCFVDDG-----NGDTVELIN 211
Db 204 --HLVGD-----TCISQCVKGC---HEEFMDGQVYMNRSQCYCPMWSRKRPQNV--N 249
QY 212 ADGCAIDKYLIN-----NLEPPTDLMAQGEAHVTKYADRSQLFYQCCQISTIKKEPNSECV 266
Db 250 INGLINLEYVTVSFTPNISPSD-----HCKWYEDR--VLEAIRTSIGKEVFKYELL 300

QY 267 RPOCSEPGFAGVKTGGAARPAAPAAQUR-----LLKRRSA--EP 304
Db 301 --NCTQD-----IKARLIAEKPLSNHVLRKIQACEHPIGEMQMYPKLLIKNSATETEE 353
QY 305 ENIIDVTDITLLEISDNOALPYD 329
Db 354 ENLCDSLKNOEAPYKGNCKVKVD 378

RESULT 38
US-08-461-004A-61
Sequence 61, Application US/08461004A
Patent No. 6235283
GENERAL INFORMATION:
APPLICANT: COBON, Stewart Gary
APPLICANT: MOORE, Joanna Terry
APPLICANT: JOHNSON, Law Anthony York
APPLICANT: WILADSEN, Peter
APPLICANT: KEMP, David Harold
APPLICANT: SRISKANTHA, Alagacone
APPLICANT: RIDING, George Alfred
APPLICANT: RAND, Keith No. 6235283man
TITLE OF INVENTION: DNA Encoding A Cell Membrane
TITLE OF INVENTION: Glycoprotein Of A Tick Gut
NUMBER OF SEQUENCES: 71
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W.
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/461,004A
FILING DATE: 04-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/325,071
FILING DATE: 19-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/062,109
FILING DATE: 17-MAY-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/926,368
FILING DATE: 07-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/242,196
FILING DATE: 06-JUL-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU87/00401
FILING DATE: 27-NOV-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU P14912
FILING DATE: 16-OCT-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU P12570
FILING DATE: 19-JUN-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU P9196
FILING DATE: 27-NOV-1986
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
TELEPHONE: 202 672 5300
TELEFAX: 202 672 5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 61:
SEQUENCE CHARACTERISTICS:

Job time : 28 secs

```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/461,004A
  FILING DATE: 04-JUN-1995
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 08/325,071
    FILING DATE: 19-OCT-1994
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 08/062,109
    FILING DATE: 17-MAY-1993
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 07/926,368
    FILING DATE: 07-AUG-1992
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: 07/242,196
    FILING DATE: 06-JUL-1988
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: PCT/AU87/00401
    FILING DATE: 27-NOV-1987
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: AU P14912
    FILING DATE: 16-OCT-1987
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: AU P12570
    FILING DATE: 19-JUN-1987
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: AU PH9196
    FILING DATE: 27-NOV-1986
  ATTORNEY/AGENT INFORMATION:
    NAME: BENT, Stephen A.
    REGISTRATION NUMBER: 29,768
    REFERENCE/DOCKET NUMBER: 60042/152
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: 202 672 5300
    TELEFAX: 202 672 5399
    TELEX: 904136
  INFORMATION FOR SEQ ID NO: 63:
    SEQUENCE CHARACTERISTICS:
      LENGTH: 650 amino acids
      TYPE: amino acid
      TOPOLOGY: linear
  MOLECULE TYPE: protein
  US-08-461-004A-63

```

Query Match 3.8%; Score 78; DB 4; Length 650;
Best Local Similarity 23.4%; Pred. No. 5.7;
Matches 62; Conservative 26; Mismatches 95; Indels 82; Gaps 15;

```

QY 115 RAYVQC---FYMEADKTVSAQIEVSEITTAFOYIV-PMFVCRYE---IIDGPTGQPV 167
      :| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 227 KAYCTCPRGTVAEIDGTCISHTVSCFAEQKOTCRPTEDCHVHKGTVLCCEPMNQ-- 284
      :| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 168 QFALIGQPVYHKWTCDSFTVDTFCVAVHSCFVDDG-----NSDTVELTN 211
      :| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 285 --HLVGD-----TCISDCYDKKC--HEEPMDCGYTNMNSCYCPWKSRRKPGPNV--N 330
      :| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 212 ADGCALDKYLIN-----NLEYPTDLMAQGEAHVRYKADRSQLFYQCOISITIKEPNSECV 266
      :| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 331 INGLNMEYTYVSFTNISFSD-----HCKWYEDR--VLEAIRTSIGKEVFKVEIL 381
      :| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 267 RPOCSEPOGAGVVTGGAAPAAAOIR-----LKKRSA---EP 304
      :| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 382 --NCTOD-----IKARLIAEKPLSNHYRLQACEHPHIGWCMMYPKILLIKNSATEIEE 434
      :| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 305 ENIDVPRDINTLEISDDNOALPYD 329
      :| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 435 ENLCDSLKNOEAAVKGONKCVKVD 459
      :| | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Search completed: March 24, 2003, 08:52:14

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 24, 2003, 08:52:19 ; Search time 18 Seconds

(without alignments)
1149.438 Million cell updates/sec

Title: US-10-054-562A-4

Perfect score: 2026

Sequence: 1 MMIRLAFCTLLALSYSTIP.....ALIAAYITTSFKRENOKA 387

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 221153 seqs, 53462247 residues

Total number of hits satisfying chosen parameters: 221153

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```
1: /cgn2_6/ptodata/1/pubppaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubppaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubppaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubppaa/US07_NEW_PUB.pep:*
5: /cgn2_6/ptodata/1/pubppaa/US07_PUBCOMB.pep:*
6: /cgn2_6/ptodata/1/pubppaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubppaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubppaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubppaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubppaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubppaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubppaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubppaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubppaa/US60_PUBCOMB.pep:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2026	100.0	387	9	US-10-054-562A-4
2	1086.5	53.6	245	9	US-10-054-562A-17
3	776.5	38.3	271	9	US-10-054-562A-9
4	92	4.5	896	10	US-09-923-563A-1
5	87.5	4.3	474	10	US-09-833-147-28
6	87	4.3	473	10	US-09-833-147-29
7	85.5	4.2	1118	9	US-09-909-567B-42
8	85	4.2	2069	9	US-09-738-626-4320
9	84.5	4.2	475	10	US-09-729-454-1
10	83	4.1	1222	10	US-09-137-531-15
11	83	4.1	1252	10	US-09-137-531-9
12	81.5	4.0	1948	9	US-09-808-602-55
13	80	3.9	293	9	US-09-738-626-5474
14	80	3.9	26926	9	US-09-759-508B-2
15	79.5	3.9	478	10	US-09-835-147-8
16	78.5	3.9	299	9	US-09-974-298-132
17	78.5	3.9	299	10	US-09-799-848-1
18	78.5	3.9	434	10	US-09-881-457A-4
19	77.5	3.8	646	9	US-10-121-746-10

20	77	3.8	1706	10	US-09-864-761-46862	Sequence 46862, A
21	77	3.8	4563	9	US-09-870-759-128	Sequence 128, App
22	77	3.8	4563	9	US-09-802-640-92	Sequence 32, App1
23	76.5	3.8	295	9	US-09-738-626-4782	Sequence 4782, App
24	76.5	3.8	1184	9	US-09-992-598-124	Sequence 124, App
25	76.5	3.8	1184	9	US-09-989-293A-124	Sequence 124, App
26	76.5	3.8	1184	9	US-09-989-735-124	Sequence 124, App
27	76.5	3.8	1184	9	US-09-980-444-124	Sequence 124, App
28	76.5	3.8	1184	9	US-09-989-730-124	Sequence 124, App
29	76.5	3.8	1184	9	US-09-930-436-124	Sequence 124, App
30	76.5	3.8	1184	9	US-09-991-181-124	Sequence 124, App
31	76.5	3.8	1184	9	US-09-993-687-124	Sequence 124, App
32	76.5	3.8	1184	9	US-09-989-734-124	Sequence 124, App
33	76.5	3.8	1184	9	US-10-028-072-412	Sequence 412, App
34	76.5	3.8	1184	9	US-09-997-653-124	Sequence 124, App
35	76.5	3.8	1184	9	US-09-993-667-124	Sequence 124, App
36	76.5	3.8	1184	9	US-10-121-049-412	Sequence 412, App
37	76.5	3.8	1184	9	US-10-123-904-412	Sequence 412, App
38	76.5	3.8	1184	9	US-10-140-470-412	Sequence 412, App
39	76.5	3.8	1184	9	US-09-980-438-124	Sequence 124, App
40	76.5	3.8	1184	9	US-09-990-562-124	Sequence 124, App
41	76.5	3.8	1184	9	US-09-997-428-124	Sequence 124, App
42	76.5	3.8	1184	9	US-09-997-666-124	Sequence 124, App
43	76.5	3.8	1184	9	US-10-175-746-412	Sequence 412, App
44	76.5	3.8	1184	9	US-10-176-918-412	Sequence 412, App
45	76.5	3.8	1184	9	US-10-176-921-412	Sequence 412, App

ALIGNMENTS

```
RESULT 1
US-10-054-562A-4
; Sequence 4, Application US/10054562A
; Patent No. US20020165375A1
; GENERAL INFORMATION:
; APPLICANT: Chandrasekhar, Ramaswamy
; APPLICANT: Morales, Tony H.
; TITLE OR INVENTION: PARASITIC HELMINTH CUTICLIN PROTEINS, NUCLEIC ACID MOLECULES,
; FILE REFERENCE: HW-8-2
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: 09/812,642
; PRIOR FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: 09/323,427
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 387
; TYPE: PRT
; ORGANISM: Dirofilaria immitis
US-10-054-562A-4

Query Match          100.0%; Score 2026; DB 9; Length 387;
Best Local Similarity 100.0%; Pred. No. 1.1e-201; Indels 0; Gaps 0;
Matches 387; Conservative 0; Mismatches 0;

QY 1 MMIRLAFCTLLALSYSTIPVDNGVEGEPEIECPSTITINFNRNAFEHVVYKGLYDQ 60
1 MMIRLAFCTLLALSYSTIPVDNGVEGEPEIECPSTITINFNRNAFEHVVYKGLYDQ 60
1 MMIRLAFCTLLALSYSTIPVDNGVEGEPEIECPSTITINFNRNAFEHVVYKGLYDQ 60
QY 61 EGCNNDGSGROVAGISLPFDSNVARTRSLNPRGIFVTTVVVISFHPLEVTYKVDRAVRVQ 120
61 EGCNNDGSGROVAGISLPFDSNVARTRSLNPRGIFVTTVVVISFHPLEVTYKVDRAVRVQ 120
Db 61 EGCNNDGSGROVAGISLPFDSNVARTRSLNPRGIFVTTVVVISFHPLEVTYKVDRAVRVQ 120
QY 121 CFVWEAKTYSAGQLEVESEITATPOTQIVPMVCVCEIILDGSPGQPVQAFATIGPVRVHKW 180
121 CFVWEAKTYSAGQLEVESEITATPOTQIVPMVCVCEIILDGSPGQPVQAFATIGPVRVHKW 180
Db 121 CFVWEAKTYSAGQLEVESEITATPOTQIVPMVCVCEIILDGSPGQPVQAFATIGPVRVHKW 180
QY 181 TCDEFYDTFCAYVWSCFVDDGNGDTVEIINADSCALDKYLLNNLEKPTDLMAGOEAHVY 240
181 TCDEFYDTFCAYVWSCFVDDGNGDTVEIINADSCALDKYLLNNLEKPTDLMAGOEAHVY 240
```

```
DB 181 TCSEVDFPCAVHSCFVDDGNGDVYELLNADGALDKYLNNLEYPTDLMAQGEAHY 240
;
QY 241 KYADRSOLFQYQOISITIKEPNSCVRPQCEPQGFAGVKTGAAKPAQAOLRLKKR 300
;
DB 241 KYADRSOLFQYQOISITIKEPNSCVRPQCEPQGFAGVKTGAAKPAQAOLRLKKR 300
;
QY 301 SAEPEINIDRTDINTLEISDDNOLPVDLRHALLQHNQPIYLAOVGICMSPFGES 360
;
DB 301 SAEPEINIDRTDINTLEISDDNOLPVDLRHALLQHNQPIYLAOVGICMSPFGES 360
;
QY 361 MEMGLSIALIAVITITSEKFRPNOKA 387
;
DB 361 MEMGLSIALIAVITITSEKFRPNOKA 387
;
RESULT 2
US-10-054-562a-17
; Sequence 17, Application US/10054562A
; Patent No. US20020165375A1
; GENERAL INFORMATION:
; APPLICANT: Chandrashekar, Ramaswamy
; APPLICANT: Morales, Tony H.
; TITLE OF INVENTION: PARASITIC HELMINTH CUTICLIN PROTEINS, NUCLEIC ACID MOLECULES, AND
; FILE REFERENCE: HM-8-2
; CURRENT APPLICATION NUMBER: US/10/054,562A
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: 09/812,642
; PRIOR FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: 09/323,427
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 17
; LENGTH: 245
; TYPE: PR
; ORGANISM: Brugia malayi
US-10-054-562a-17
;
Query Match 53.6%; Score 1086.5; DB 9; Length 245;
Best Local Similarity 80.6%; Pred. No. 1.4e-104;
Matches 195; Conservative 24; Mismatches 22; Indels 1; Gaps 1;
;
QY 3 IRILACTLLIALST-SIVDNKVEGEPEIECGPTSIITINENRMAFEGHYVKGILYDQ 61
;
DB 4 MOICSLSYMILASINAIIDNGVESEPEIECGPTSIITVFNENRNPFECHVYAKGILYDQ 63
;
QY 62 GCRNDEGGRQVAGISLPFSCNVAARTRSINPRGIEFTTVVISFHPDLFTYKVDRAVYOC 121
;
DB 64 DCRSDGGRQVAGISLPFSCNVAARTRSINPRGIEFTTVVISFHPDLFTYKVDRAVYOC 123
;
QY 122 FYMEADKTVSAOIEVSEITTAFOQIIVMPVCRYEILIDGPGTGOVQFAIIGQPVYHKMT 181
;
DB 124 FYMEADKTVSTOIEVSEMTTFATQIIVMPVCRYEILIDGPGTGOVQYANIGQPVYHKMT 183
;
QY 182 CSEVDFPCAVHSCFVDDGNGDVYELLNADGALDKYLNNLEYPTDLMAQGEAHYK 241
;
DB 184 CSEVDFPCAVHSCFVDDGNGDVYELLNADGALDKYLNNLEYPTDLMAQGEAHYK 243
;
QY 242 YA 243
;
DB 244 YA 245
;
RESULT 3
US-10-054-562a-9
; Sequence 9, Application US/10054562A
; Patent No. US20020165375A1
; GENERAL INFORMATION:
; APPLICANT: Chandrashekar, Ramaswamy
; APPLICANT: Morales, Tony H.
; TITLE OF INVENTION: PARASITIC HELMINTH CUTICLIN PROTEINS, NUCLEIC ACID MOLECULES, AND
; FILE REFERENCE: HM-8-2
```

```
;
; CURRENT APPLICATION NUMBER: US/10/054,562A
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: 09/812,642
; PRIOR FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: 09/323,427
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
; LENGTH: 271
; TYPE: PR
; ORGANISM: Dirofilaria immitis
US-10-054-562a-9
;
Query Match 38.3%; Score 776.5; DB 9; Length 271;
Best Local Similarity 72.4%; Pred. No. 2.2e-72;
Matches 144; Conservative 24; Mismatches 22; Indels 9; Gaps 2;
;
QY 134 IEVSEITTAFOQIIVMPVCRYEILIDGPGTGOVQFAIIGQPVYHKMTCDSEVDFPCAV 193
;
DB 15 LEVSEMTTAFQIIVMPVCRYEILIDGPGTGOVQFAIIGQPVYHKMTCDSEVDFPCAV 193
;
QY 194 VHSCEVDDGNGDVYELLNADGALDKYLNNLEYPTDLMAQGEAHYKYADRSOLFQYQO 253
;
DB 75 VHSCEVDDGNGDVYELLNADGALDKYLNNLEYPTDLMAQGEAHYKYADRSOLFQYQO 253
;
QY 254 ISITIKEPNSCVRPQCEPQGFAGVKTG-----GAAKPAQAOL---RLKKRSAP 304
;
DB 135 ISITIKEPNSCVRPQCEPQGFAGVKTG-----GAAKPAQAOL---RLKKRSAP 304
;
QY 305 ENIIDVTRDINTLEISDDN 323
;
DB 195 DNTVDVSTGSGTVDITEEN 213
;
RESULT 4
US-09-923-563a-1
; Sequence 1, Application US/09923563A
; Patent No. US20020115141A1
; GENERAL INFORMATION:
; APPLICANT: Wambach, Wolfgang
; APPLICANT: Bayer AG
; TITLE OF INVENTION: Method for Improved Production of Cyanophycin and Secondary Pr
; FILE REFERENCE: Lea 34650
; CURRENT APPLICATION NUMBER: US/09/923,563A
; CURRENT FILING DATE: 2001-08-07
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 896
; TYPE: PR
; ORGANISM: Synechococcus elongatus
US-09-923-563a-1
;
Query Match 4.5%; Score 92; DB 10; Length 896;
Best Local Similarity 22.0%; Pred. No. 1.3;
Matches 91; Conservative 44; Mismatches 155; Indels 124; Gaps 22;
;
QY 36 TSTITINFRNMAFEGHYVKGILYDQEGCRNDEGGRQVAGISLPFSCNVAARTRSINPRGI 95
;
DB 356 TRIEVNHDTWTLEROGITLNTVLPQ-----ETCYLRATANLSTGSI 398
;
QY 96 EFTTVVISFHPDLFTYKVDRAVYOCFYMEADKTVSAOIE-----VSEITTA--EQ 144
;
DB 399 AIDRDEI--HPEWNWICQRAARIIGDLAIDVSPDISQSLKSGVGYEVNNAAPGR 456
;
QY 145 TQIVPM-----PVCRYEILIDGPGTGOVQFAIIGQPVYHKMTCDSEVDFPCAVH 195
;
DB 457 MTNPSQGIARVAVBPVLNMLPPPGPCRIPI-FAITG-----TNGKTTT--RLAH 506
;
QY 196 SCFVDDGNGDVYELLNADGALDKYLNN-----LEYPTDLMAQGEAHYKYA 243
```

```
Db 507 IC---KQGGVGTGTTTGITGIVLVEKGDPTGPOSAQLLQDPT-----VELAVLETA 558
QY 244 D----RSOL-FYQCOIST-----TIKE-PNSECVRPQCSPOGGAVTG 282
Db 559 RGLIRSGIIGDHCIVGVWADHLGLDIDVEQLADIKAVVESAMNGYAVLN-- 616
QY 283 GAAKPAQAOLRLKKR-----SAEPENIDVRDINT-----LEISDNQALP 327
Db 617 --ADPLVAAAROVKAQAVAFSMDPHNP-IROHIQGGIAAYENGLISILKGDWTLR 673
QY 328 VDLHRALLQHNQPVIIAA-----VQNGICSPFGSMFGLSTALIAVIT 376
Db 674 IE-----QAEVNPITLIGARASFMIANAALASIAFA--QGISTEIRALTT 718
```

```
RESULT 5
US-09-835-147-28
; Sequence 28, Application US/09835147
; Patent No. US2002000227A1
; GENERAL INFORMATION:
; APPLICANT: Maliszewski, Charles R.
; APPLICANT: Gayle III, Richard B.
; APPLICANT: Price, Virginia L.
; APPLICANT: Gimpel, Steven D.
; TITLE OF INVENTION: Inhibitors of Platelet Activation and Recruitment
; FILE REFERENCE: 2879-US
; CURRENT APPLICATION NUMBER: US/09/835,147
; CURRENT FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: US 60/104,585
; PRIOR FILING DATE: 1998-10-16
; PRIOR APPLICATION NUMBER: US 60/107,466
; PRIOR FILING DATE: 1998-11-06
; PRIOR APPLICATION NUMBER: US 60/149,010
; PRIOR FILING DATE: 1999-08-13
; PRIOR APPLICATION NUMBER: PCT/US99/22955
; PRIOR FILING DATE: 1999-10-13
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 28
; LENGTH: 474
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Fusion
; OTHER INFORMATION: construct of human CD39
US-09-835-147-28
```

```
Query Match 4.3% Score 87.5; DB 10; Length 474;
Best Local Similarity 20.9%; Pred. No. 1.4;
Matches 55; Conservative 40; Mismatches 75; Indels 93; Gaps 13;

QY 113 VDRAVRVOCFMYE-ADKTVSAQIEVSEITTAFOQIVPMPCRYEILDDGPGPQVQFAI 171
Db 5 IDRMQLSCIALSLALVTNSSTKKTQLTSSQNKALPENKYGIVLDAGSSHTSLYT-- 62
QY 172 IGQPVYHKWTCDESEVDFCAVH-----SCFVDDGN----- 203
Db 63 -----YKMPAREKEN-DT--GVVHVEECRVKPGISKRFQKVNELGIYLTDCMERAREV 113
QY 204 -----GDT--VELINADGALDKYLLNLE-----YPTD-----LMAGQEAHV 239
Db 114 IPRSOHETPVYLGATAGRLLRMESEELADRVLDIVERSLSNYPDFQGARITIGOEGB 173
QY 239 VY-----KYADRSOLFQCOISITIKENPNSCVRPQCSPEPGFAGVKTGGAAPK 288
Db 174 AVGTTITVYLKGFQSKTRWF-----SIVPYETNNQ-----ETFGALDLGGASTQV 219
QY 289 AAAAQLRLKKRSAPENIIDVR 311
Db 220 TFPVQNGTIE-----SPDNALQFR 238
```

```
RESULT 6
US-09-835-147-29
; Sequence 29, Application US/09835147
; Patent No. US2002000227A1
; GENERAL INFORMATION:
; APPLICANT: Maliszewski, Charles R.
; APPLICANT: Gayle III, Richard B.
; APPLICANT: Price, Virginia L.
; APPLICANT: Gimpel, Steven D.
; TITLE OF INVENTION: Inhibitors of Platelet Activation and Recruitment
; FILE REFERENCE: 2879-US
; CURRENT APPLICATION NUMBER: US/09/835,147
; CURRENT FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: US 60/104,585
; PRIOR FILING DATE: 1998-10-16
; PRIOR APPLICATION NUMBER: US 60/107,466
; PRIOR FILING DATE: 1998-11-06
; PRIOR APPLICATION NUMBER: US 60/149,010
; PRIOR FILING DATE: 1999-08-13
; PRIOR APPLICATION NUMBER: PCT/US99/22955
; PRIOR FILING DATE: 1999-10-13
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 29
; LENGTH: 473
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Fusion
; OTHER INFORMATION: construct of human CD39
US-09-835-147-29
```

```
Query Match 4.3% Score 87; DB 10; Length 473;
Best Local Similarity 19.5%; Pred. No. 1.6;
Matches 51; Conservative 42; Mismatches 77; Indels 92; Gaps 12;

QY 113 VDRAVRVOCFMYEADKTVSAQIEVSEITTAFOQIVPMPCRYEILDDGPGPQVQFAI 172
Db 5 IDRMQLSCIALSLALVTNSSTKKTQLTSSQNKALPENKYGIVLDAGSSHTSLYT-- 61
QY 173 GQPVYHKWTCDESEVDFCAVH-----SCFVDDGN----- 203
Db 62 -----YKMPAREKEN-DT--GVVHVEECRVKPGISKRFQKVNELGIYLTDCMERAREV 113
QY 204 -----GDT--VELINADGALDKYLLNLE-----YPTD-----LMAGQEAHV 239
Db 114 IPRSOHETPVYLGATAGRLLRMESEELADRVLDIVERSLSNYPDFQGARITIGOEGB 173
QY 240 Y-----KYADRSOLFQCOISITIKENPNSCVRPQCSPEPGFAGVKTGGAAPK 289
Db 174 YGTTITVYLKGFQSKTRWF-----SIVPYETNNQ-----ETFGALDLGGASTQV 219
QY 290 AAAQLRLKKRSAPENIIDVR 311
Db 220 FVPVQNGTIE-----SPDNALQFR 237

RESULT 7
US-09-909-567B-42
; Sequence 42, Application US/09909567B
; Publication No. US20030022257A1
; GENERAL INFORMATION:
; APPLICANT: Macina, Roberto A.
; APPLICANT: Nair, Manoj
; APPLICANT: Chen, Selju
; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes
; FILE REFERENCE: DEX-0214
; CURRENT APPLICATION NUMBER: US/09/909,567B
; CURRENT FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: 60/219,834
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: PatentIn version 3.1
```

```
; SEQ ID NO 42
; LENGTH: 1118
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-909-567B-42

Query Match
Best Local Similarity 25.6%; Score 85.5; DB 9; Length 1118;
Matches 52; Conservative 27; Mismatches 81; Indels 43; Gaps 12;

15 LSTSIPLVNDNGVEGEPEI--ECGPTSIITINFTNNAFEG-----HVVYKGLYDQSGCR 64
Db 348 LBSVLYANKAAAFHKSRKSTYKPTPLPIKVKVKNVKEGHHFKSSVSTIVSESD---R 403
QY 65 NDEGGRVAGISLPFDSGN---VARTSLNPRGIF-----VTVVVISFHPPLFTVKV-- 113
Db 404 SSKG--QIIGNFQAFDEDTGLPAHARYVKLEDDNMISVDSVTSSEIKLAKLPDESEHYQ 461
QY 114 DRAIRVQCFMEAD---KIVSAQ--IEVSEITTAFOQTQVPM-PVCR-----YEILD 159
Db 462 NGTYTVKIVAISEDYPRKTTGTVLINVEDINDNCPITLIEPVOTICHDAEYVNVTAEDLD 521
QY 160 GCPTGQPVQFAIGOP--VYHKW 180
Db 522 GHNSGPFSTSVIDKPRGMAEKW 544

RESULT 8
US-09-738-626-4320
; Sequence 4320, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 4320
; LENGTH: 2069
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-4320

Query Match
Best Local Similarity 4.2%; Score 85; DB 9; Length 2069;
Matches 73; Conservative 50; Mismatches 116; Indels 124; Gaps 17;

15 LSTSIPLVDN-----GVEGEPEIECGPTSIITINFTNNAFEGHYVVKGLYDQSGC 63
Db 1431 VEFVVRPDDHGKSMWYLNDFEGIEHLPKVELRWMLNLGIGNGQRRLGCGFEVTSLEFNV--C 1488
QY 64 R-----NDEGGRVAGISLPF-----DSCNVAARTSLNPRGIFVTVVVISFHPPLF 109
Db 1489 RHCGHLDSEAGANSRWDHRPMCPHRYEKEDVTYSPALGRTLTKQGVLMLLP----- 1539
QY 110 VTKVDRAYVQCFYMEADKTVSAQIEVSEITTAFOQTQVPMFVCRKEIILDGPTGQPVQF 169
```

```
Db 1540 -----EYFGSEADSNVVTSL-IAAIKILGFR-----EVLGDP----- 1570
QY 170 AIIIGPYVYHKWTCDSSENVDFECVAVHSCFVDDGNGDTVELLNADGALDKYLLNNLEYPT 229
Db 1571 -----DHLD-----VTSVOVPRTSGDG-----ALDALLHL----- 1595
QY 230 DLMAQGEAHYKYKADRSOL-----FYQCOISITIKPEPNSCYAPQCSFQFG 277
Db 1596 DQVPGGTGYLNGFADPTKVELLSRAMERYSRGQCYDERLACP--ECLLPYRTAT--- 1650
QY 278 AVTGGAAKPAAPAAQRLRLKRSABENIIDVT--DINTLEISDNO-ALPVDLRHRA 334
Db 1651 LLOTSTRAAAEKA-----LRAILLNNSRBEETIDLSAVPDWTFLEKREPNLTGSOLELRFV 1706
QY 335 LLD 337
Db 1707 MLR 1709

RESULT 9
US-09-729-454-1
; Sequence 1, Application US/09729454
; Patent No. US20020137038A1
; GENERAL INFORMATION:
; APPLICANT: Lasek, Amy W.
; APPLICANT: Krasnow, Randi E.
; APPLICANT: Baughn, Mariah R.
; TITLE OF INVENTION: INTESTINAL PROTEINS
; FILE REFERENCE: PC-0028 CIP
; CURRENT APPLICATION NUMBER: US/09/729,454
; CURRENT FILING DATE: 2000-12-04
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PERL Program
; SEQ ID NO 1
; LENGTH: 475
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. US20020137038A1 3229449CD1
US-09-729-454-1

Query Match
Best Local Similarity 4.2%; Score 84.5; DB 10; Length 475;
Matches 62; Conservative 32; Mismatches 92; Indels 143; Gaps 12;

QY 87 TBSLNPRGIFVTVVVISFHPLEFVTKVDARYVQCFYMEADKTVSAQIEVSEITTAFOQT 146
Db 50 TSLFPR-----TPLISLKPLETE-----ELRIKETIEKLDQ 82
QY 147 IYPMVPCRYEILDGPTGQPVQFAIGOPVYHKWTCDSF-----VDFECVAVHSC 197
Db 83 IPP-----RPTHYNTTTSATHSTATILNPRDYC----- 112
QY 198 FVDDGNGDTVELLNADGALDKYLLNNLEYPTD-----LMAQGEAHYKYKADRS- 246
Db 113 -----RGDQHLI---LEVRDHLGRKQYGDGFLRARSSPALMAGAGKTYDFENGTY 163
QY 247 -----QLFYQCOISITIT-----KEPNSC----- 265
Db 164 LVSFTLFWEGVSLSLILIHSEGVSAIWSARNQGYDRVIFTGOFVNGTSOVHSECGLIL 223
QY 266 -----VRPQ-----CSEPGFCAVKTGGAAPAAQRLRLKRSABENI-ID 309
Db 224 NNNAELCYLDNRDDEGFYCVAPQNHPCALITHMSKMKKVSYLKQEKSLFERNSGYE 283
QY 310 VRTDINTLEISDNOALPYDLRHRALLH 338
Db 284 IMEKFWTISVSKNTLKSVDLHESGKLOH 312

RESULT 10
```


US-09-137-531-15
; Sequence 15, Application US/09137531
; Patent No. US20020048816A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Expression of surface layer proteins
; NUMBER OF SEQUENCES: 25
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/137,531
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/682,517
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1222 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-137-531-15

Query Match 4.1%; Score 83; DB 10; Length 1222;
Best Local Similarity 19.8%; Pred. No. 17;
Matches 72; Conservative 50; Mismatches 133; Indels 108; Gaps 17;

QY 66 DEGRVAGISLPDSCNVART-----RSLNPGIEVTT-----TVVISFHP 107
DB 676 NQSKKVTGTSIKRATYTYNTGANDIKVDNQVISPNRSYTVTEATLSSTGTVITPAKN 735
QY 108 LEVTKVD-----RAYRVOCFYMEADKTVSAOIEVSEITTAFOQIV 148
DB 736 LEVTSVQDKTTAVKVIATGIAVNTDGDYATTA--KEATATFTATNEVPNSYTGVAIO-- 791
QY 149 PMPVCRYEILDGPGTPGVQVFAIGO--PVYHKWTCDETVDFCAVHSCFVDGNGDPTV 207
DB 792 -----FNTADSGSNSIMFA--GKNPVKAGV-----SGKTY 822
QY 208 EILNADG-----CALDKYLLNLEYPDLMAGQEAHVKYKADRSQLYQCQISITTKEP 261
DB 823 KYFGANGNEVFEAAMEALL--TQYATE--GQKVTISYVNDGDTVFKV--ISAVNS 873
QY 262 NSECVRPOCSPEQGFAGVKTGAAPAAQAQLRLKKRSEPENIIDVRDINTLEIS- 320
DB 874 STEAIKPYVA--PTTPAAPTTGALTLLTPAG-----GLVDLTATNTLGLISL 917
QY 321 ---DDN-QALPVDLRHRLALOHNGQPYLLAIVQNGICSPFGSMFGLSIALIAAVIT 376
DB 918 ADADLNVSATVTDATATSLKNSANNSLSLTLVETGANTGVATTVQAG-TLSSLTACTLT 976
QY 377 ISF 379
DB 977 VTY 979

RESULT 11
US-09-137-531-9
; Sequence 9, Application US/09137531
; Patent No. US20020048816A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Expression of surface layer proteins
; NUMBER OF SEQUENCES: 25
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/137,531
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/682,517
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1252 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-137-531-9

Query Match 4.1%; Score 83; DB 10; Length 1252;
Best Local Similarity 19.8%; Pred. No. 18;
Matches 72; Conservative 50; Mismatches 133; Indels 108; Gaps 17;

QY 66 DEGRVAGISLPDSCNVART-----RSLNPGIEVTT-----TVVISFHP 107
DB 706 NQSKKVTGTSIKRATYTYNTGANDIKVDNQVISPNRSYTVTEATLSSTGTVITPAKN 765
QY 108 LEVTKVD-----RAYRVOCFYMEADKTVSAOIEVSEITTAFOQIV 148
DB 766 LEVTSVQDKTTAVKVIATGIAVNTDGDYATTA--KEATATFTATNEVPNSYTGVAIO-- 821
QY 149 PMPVCRYEILDGPGTPGVQVFAIGO--PVYHKWTCDETVDFCAVHSCFVDGNGDPTV 207
DB 822 -----FNTADSGSNSIMFA--GKNPVKAGV-----SGKTY 852
QY 208 EILNADG-----CALDKYLLNLEYPDLMAGQEAHVKYKADRSQLYQCQISITTKEP 261
DB 853 KYFGANGNEVFEAAMEALL--TQYATE--GQKVTISYVNDGDTVFKV--ISAVNS 903
QY 262 NSECVRPOCSPEQGFAGVKTGAAPAAQAQLRLKKRSEPENIIDVRDINTLEIS- 320
DB 904 STEAIKPYVA--PTTPAAPTTGALTLLTPAG-----GLVDLTATNTLGLISL 947
QY 321 ---DDN-QALPVDLRHRLALOHNGQPYLLAIVQNGICSPFGSMFGLSIALIAAVIT 376
DB 948 ADADLNVSATVTDATATSLKNSANNSLSLTLVETGANTGVATTVQAG-TLSSLTACTLT 1006
QY 377 ISF 379
DB 1007 VTY 1009

RESULT 12
US-09-808-602-55
; Sequence 55, Application US/09808602
; Patent No. US20020155115A1
; GENERAL INFORMATION:
; APPLICANT: Vernet, Corine A
; APPLICANT: Fernandes, Elma
; APPLICANT: Shimkets, Richard A
; APPLICANT: Herriman, John L
; APPLICANT: Majumder, Kumud
; APPLICANT: Mishra, Vishnu
; APPLICANT: Mezes, Peter S
; TITLE OF INVENTION: No. US20020155115a1 Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 15966-697 CIP
; CURRENT APPLICATION NUMBER: US/09/808,602
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 09/800,198
; PRIOR FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: 60/186,596
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 55
; LENGTH: 1948

```
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-808-602-55

Query Match
Best Local Similarity 4.0%; Score 81.5; DB 9; Length 1948;
Matches 84; Conservative 47; Mismatches 137; Indels 123; Gaps 20;

QY 27 GEPET-DCGPTSTININTRAHFGHYV-----KGLY----DDEGCRNDEGHQ-- 71
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 336 GTPWNTNTATSTITWDSGNPDVSYVIEYKSKSODGPQIKEDITTTTYSIGLSPN 395
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 72 -----VAGISL-----PFDSCNVART-----RSINPGIVTPTTVVISFHLVTK 112
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 396 SEYIHWSAVNSIGQGPSESVYRTGEQAPASPRVQARMISATMIYQMEEPVENG 455
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 113 VDRAYRVOCFYMEADKTV---SQAIEVSEITTA---FQIQIYPMPCRYEILDGPTGQ 165
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 456 LIRGFRVY-YTMEPEHFGVMNQKHNDLSLTITVGSLLDEBTYVRLAFTSVGDPLSD 514
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 166 PVQF---AIIQGVYHAKTCDSTVTFC-----AVV--HSCFVDDGNGDTVELLNA 212
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 515 PIVQKTQGVGPQPMNLRABARSETSTLSMSPROBSIIRKYEELFREGDHGREV----- 569
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 213 DGCALD---KYLINMLEYPTDLMAQGEAHVYKYADRSQLFYQOISITIKENSECVRQ 269
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 570 -GRFEDPTTSYVEDLKPNT-----YAFRLAARS----- 598
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 270 CSEPOFGAVKGTGAAPAAQAQLRLKRSAPENI---IDVTDINTLEISDDNOALP 327
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 599 ---PQIGAF-----TPVVRQRTLOSQKPSAPQDVKCVSYR-----STAIL 636
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 328 VDLRHRALLHONGQPVTLAAVQNGICMSPFG 358
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 637 VSMRPPPTTHNG-----ALVGYSVRYRPLG 662
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 13
US-09-738-626-5474
; Sequence 5474, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: NAKAGUCHI, HIROSHI
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: SEIKO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738, 626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 5474
; LENGTH: 293
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-5474

Query Match
Best Local Similarity 3.9%; Score 80; DB 9; Length 293;
Matches 31; Conservative 10; Mismatches 47; Indels 16; Gaps 4;
```

```
QY 74 GILSPEDSCNVARTSRSLNPGIIVTPTTVVISFHLVTKVDRAYRVOCFYMEADKTVSAQ 133
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 137 GKSLP-----AVMTTKSLQPSGVDPISDAMNIQOOOFMAVAMNRQ-----AAEAQKAAQ 186
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 134 IEVSEITTAFOQIYPMPCRYEILDGSPFGQVQFAITIQPVY 177
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 187 -----ATQADAPVAPQPVAMNQOQFQ-APVPQGOQPAVAGAPVY 224
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 14
US-09-759-508B-2
; Sequence 2, Application US/09759508B
; Publication No. US20020182599A1
; GENERAL INFORMATION:
; APPLICANT: Fishman, Mark C.
; TITLE OF INVENTION: Methods for Diagnosing and Treating Heart Disease
; FILE REFERENCE: 00786/381002
; CURRENT APPLICATION NUMBER: US/09/759, 508B
; CURRENT FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: US 60/175,787
; PRIOR FILING DATE: 2000-01-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 26926
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-759-508B-2

Query Match
Best Local Similarity 3.9%; Score 80; DB 9; Length 26926;
Matches 51; Conservative 46; Mismatches 116; Indels 62; Gaps 10;

QY 107 PLEFVKVDRAYRVOCFYMEADKTVSAQIEVSEITTAFOQIYPMPCRYEILDGPTGQ 166
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1036 PYRTTK-----PVQKLVREGSVYEGQVGNKPKPVYTK-----KSGVP 1075
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 167 VQFAIIGOPVYHAKTCDSETVDTPFCVAVHSCFVDDGNGDTVELLNADG-CALDKYLLNLI 225
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1076 LITGYRYKVSYNQTOGECK-----LVISMTFADAGEYTIYVRNKHGFTSASASLLEEA 1129
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 226 EYPTDLMAQGEAHVYKABRSQLFYQCQISITIKENSECVRQCSPEQFGAVKTGGA 285
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1130 DYEL-LMKSQQ-----EMLYQOTVATVQVEPEVETAP-----GFVYSEKEY 1172
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 286 AKPAAAQAQLRLKRSAPENIIDVTDINTLEISDDNOALPVDLRH-----ALLQH 338
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1173 EKEQA-----LIRKKAQDTVYVRYVEDQEHISFEERLIKEIYRIKTTLEELLEE 1227
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 339 NGQPVTLA-----AVQNGICMSPFGSMFGLSI 367
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1228 DGEKKAVIDISEAVESGFDLRKNRYRLBGMGV 1262
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 15
US-09-835-147-8
; Sequence 8, Application US/09835147
; Patent No. US20020002277A1
; GENERAL INFORMATION:
; APPLICANT: Maliszewski, Charles R.
; APPLICANT: Gayle III, Richard B.
; APPLICANT: Price, Virginia L.
; APPLICANT: Gimpel, Steven D.
; TITLE OF INVENTION: Inhibitors of Platelet Activation and Recruitment
; FILE REFERENCE: 2879-US
; CURRENT APPLICATION NUMBER: US/09/835, 147
; CURRENT FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: US 60/104,585
; PRIOR FILING DATE: 1998-10-16
; PRIOR APPLICATION NUMBER: US 60/107,466
; PRIOR FILING DATE: 1998-11-06
; PRIOR APPLICATION NUMBER: US 60/149,010
```

189 TFCVVHSCFVDDGNGDTVEILNADGALD-----KYLNNLEYPTDLMAGQEAHVYKYA 243

```
Db 79 LFAKMEJDTEV-----GHALDVLSPNLISTSMLGWKNYNLSISMDT-----ANGLLQYA 128
QY 244 -----DSOLFYOQOISITTKERNSECVRPOQSEPGFAGVKTGGAAPK-----AAAAQLR 295
Db 129 LDQGVNATQVFEVD-----TWGMPETQARLQOQSFPGIEVTKAKADALYPVSAASICAK 183
QY 296 LKRRSAEPENIIDVTDINTLEISDDNQALPVDLRHRLLOHNGQPVILAAVONGICMS 355
Db 184 VARDQAVKKMQFVEKLODDLT-----DYGSGYFNDPKTKAKMLKEHVEFV-----227
QY 356 PFGFSNPMGLS 366
Db 228 -FGFPQFVFRFS 237

RESULT 18
US-09-881-457A-4
; Sequence 4, Application US/09881457A
; Patent No. US20020081316A1
; GENERAL INFORMATION:
; APPLICANT: Cochran, Mark D
; APPLICANT: Cook, Stephanie M
; APPLICANT: Wild, Martha A
; TITLE OF INVENTION: NO. US20020081316A1e1 Avian Herpes Virus and Uses Thereof
; FILE REFERENCE: SY01105K1OKOK
; CURRENT APPLICATION NUMBER: US/09/881,457A
; PRIOR FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: 09/426,352
; PRIOR FILING DATE: 1999-10-25
; PRIOR APPLICATION NUMBER: 08/804,372
; PRIOR FILING DATE: 1997-02-21
; PRIOR APPLICATION NUMBER: PCT/US95/10245
; PRIOR FILING DATE: 1995-08-09
; PRIOR APPLICATION NUMBER: 08/663,566
; PRIOR FILING DATE: 1996-06-13
; PRIOR APPLICATION NUMBER: 08/288,065
; PRIOR FILING DATE: 1994-08-09
; PRIOR APPLICATION NUMBER: PCT/US93/05681
; PRIOR FILING DATE: 1993-06-14
; PRIOR APPLICATION NUMBER: 08/023,610
; PRIOR FILING DATE: 1993-02-26
; PRIOR APPLICATION NUMBER: 07/898,087
; PRIOR FILING DATE: 1992-06-12
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 434
; TYPE: PRT
; ORGANISM: Infectious laryngotracheitis Virus
US-09-881-457A-4

Query Match 3.9%; Score 78.5; DB 10; Length 434;
Best Local Similarity 18.9%; Pred. No. 10;
Matches 80; Conservative 55; Mismatches 148; Indels 141; Gaps 22;

QY 45 RNAFEGHYVGLVDQECRNDGGRQVAGISLPDSCNVARTSLNRRGIFVTTTIVIS 104
Db 55 RNAADRHLEFKNAFTWTTLYLLSSPASQSTAAVTYD-----ILGRRALDALITPAVG 105
QY 105 FHPLEVRKVDRAVYOCFYME-----ADKTVSAQIEVSEITTAFOQIVMPVPCRYEIL 158
Db 106 PYNRLYTRVSRG-----COVELNLPISNDDMISAKEREK-----141
QY 159 DGGP-TGQVQFAIIT-----GQ-----PVYHK--WTQDSEYVDTFCAY-----VHSC 197
Db 142 -GGPEASVWVEYVIKGGDDGDKYCPYRKEYRRCGDVQLSECAVQSAQMAWADVVEST 200
QY 198 FVDDGNGDTVEILNADGCLDKYLL-----222
Db 201 LV-SRNAGLITFSPALASGOYLLTLKIGFAOTALTVEVNDRCIKISQNLFLPSKC 259
QY 223 -NNLEYPDLMAQGEAHVYKYADRS-----OLFYOQOISITIKPEPNSCVARQ 269
```

```
Db 260 WTEQYQYTGFE---QGEHALYPIADTNTNRHADVYRGYEDIIQWRNNLLRKRNPASADPRPD 316
QY 270 CSEPGFAGVKTGGAAPKPAAPAOQLRLKKRSAPENIIDVTDINTLEISDDNQA-----325
Db 317 -SVYQELIPATYTKKRAEGTTPAES-----SEKKAPPE---DSEDDMQA-EASGENPALPE 366
QY 326 ---LPVDLRHRLALQ-----HNGQPVILAAVO-----NGICMSPFGFSMGLSIALTAAY 373
Db 367 DDEVPEDTEHDDPNSPDYNDMPAVIPVEETKSSNAVSM-PI-FAAFVACAVALVGLL 424
QY 374 IITI 377
Db 425 VMSI 428

RESULT 19
US-10-121-746-10
; Sequence 10, Application US/10121746
; Publication No. US20030036648A1
; GENERAL INFORMATION:
; APPLICANT: Miller, Andrew P.
; APPLICANT: Curran, Mark Edward
; APPLICANT: Hu, Ping
; APPLICANT: Rutger, Marc
; APPLICANT: Wang, Jian-Wang
; TITLE OF INVENTION: NO. US20030036648A1e1 Human Potassium Channels
; FILE REFERENCE: SEQ-15P
; CURRENT APPLICATION NUMBER: US/10/121,746
; PRIOR FILING DATE: 2002-04-11
; PRIOR APPLICATION NUMBER: US/09/336,643A
; PRIOR FILING DATE: 1999-06-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/076,687
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-08-07
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/116,448
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-01-19
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: PCT/US99/03826
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-02-22
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 646
; TYPE: PRT
; ORGANISM: H. sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(646)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-10-121-746-10

Query Match 3.8%; Score 77.5; DB 9; Length 646;
Best Local Similarity 22.3%; Pred. No. 24;
Matches 43; Conservative 26; Mismatches 93; Indels 31; Gaps 7;

QY 199 VDDGNGDTVEILNADGCLDKYLLNLE-YPTDLMAQGEAHVYKYADRSOLFYOQOISIT- 256
Db 33 VKASRGDYVLVYVNSGRFFETW-KNTLDRIYDPTLLGSSEKEFFYDADSGEYFFDRDPMF 91
QY 257 -----TIKPEPNSCVRPOQSEPGFAGVKTGGAAPKPAAPAOQLRLKKRSAPEN 306
Db 92 RHYLVNFTYTGRIHCPROECIOAFDELAFFGLV-----PELVGDCCLLEYRDRKKEN 143
QY 307 IIVRTDINTLEISDDNQALPV--DLRHRALLOHNGQPVILAAVONGICMSPFGFSMWP 364
Db 144 AERLAED-EAELQAGDGPALPAGSSLRQLRRAENPHITSTAL-----VFYYVIGFF 195
QY 365 LSIALLAAVITTI 377
Db 196 IAVSIANVYETI 208

RESULT 20
US-09-864-761-46862
```

```
; Sequence 46862, Application US/09864761
; Patent No. US2002048763A1
GENERAL INFORMATION:
APPLICANT : Penn, Sharon G.
APPLICANT : Rank, David R.
APPLICANT : Hanzel, David K.
APPLICANT : Chen, Wensheng
TITLE OF INVENTION : HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE : Aemica-x-1
CURRENT FILING DATE : 2001-05-23
PRIOR APPLICATION NUMBER : US/09/864,761
PRIOR FILING DATE : 2000-02-04
PRIOR APPLICATION NUMBER : US 60/207,456
PRIOR FILING DATE : 2000-05-26
PRIOR APPLICATION NUMBER : US 09/632,366
PRIOR FILING DATE : 2000-08-03
PRIOR APPLICATION NUMBER : GB 24263.6
PRIOR FILING DATE : 2000-10-04
PRIOR APPLICATION NUMBER : US 60/236,359
PRIOR FILING DATE : 2000-09-27
PRIOR APPLICATION NUMBER : PCT/US01/00666
PRIOR FILING DATE : 2001-01-30
PRIOR APPLICATION NUMBER : PCT/US01/00667
PRIOR FILING DATE : 2001-01-30
PRIOR APPLICATION NUMBER : PCT/US01/00664
PRIOR FILING DATE : 2001-01-30
PRIOR APPLICATION NUMBER : PCT/US01/00669
PRIOR FILING DATE : 2001-01-30
PRIOR APPLICATION NUMBER : PCT/US01/00665
PRIOR FILING DATE : 2001-01-30
PRIOR APPLICATION NUMBER : PCT/US01/00668
PRIOR FILING DATE : 2001-01-30
PRIOR APPLICATION NUMBER : PCT/US01/00663
PRIOR FILING DATE : 2001-01-30
PRIOR APPLICATION NUMBER : PCT/US01/00662
PRIOR FILING DATE : 2001-01-30
PRIOR APPLICATION NUMBER : PCT/US01/00661
PRIOR FILING DATE : 2001-01-30
PRIOR APPLICATION NUMBER : PCT/US01/00670
PRIOR FILING DATE : 2001-01-30
PRIOR APPLICATION NUMBER : US 60/234,687
PRIOR FILING DATE : 2000-09-21
PRIOR APPLICATION NUMBER : US 09/608,408
PRIOR FILING DATE : 2000-06-30
PRIOR APPLICATION NUMBER : US 09/774,203
NUMBER OF SEQ ID NOS : 49117
SOFTWARE : Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 46862
LENGTH : 1706
TYPE : PRF
ORGANISM : Homo sapiens
FEATURE:
OTHER INFORMATION : MAP TO AC010872.3
OTHER INFORMATION : EXPRESSED IN ADULT LIVER, SIGNAL = 5.9
OTHER INFORMATION : EXPRESSED IN FETAL LIVER, SIGNAL = 8.1
OTHER INFORMATION : SWISSPROT HIT: P04114, EVALUOE 0.00e+00
OTHER INFORMATION : EST_HUMAN HIT: AA702484.1, EVALUOE 0.00e+00
OTHER INFORMATION : EST_HUMAN HIT: AA702484.1, EVALUOE 0.00e+00
US-09-864-761-46862
Query Match      3 8%; Score 77; DB 10; Length 1706;
Matches Local Similarity 18.7%; Pred.No.1.2e+02;
Match     81; Conservative    46; Mismatches 135; Indels 172; Gaps   20;
```

QY	DB	Sequence	Score	DB	Length	Matches	Indels	Gaps
QY	305	GSAYQAMILGVDSKNIFNFVKVSOEGKLKLSNDMMGSA	3.8%;	DB 9;	Length 4563;			
QY	96	FTTTTAVISFHPLEFVKRVKADRAVROCFYMBADKTVSAQIEVSEFTTAFOQIIVMPVCRY	18.7%;	Pred. No. 5.4e+02;				
DB	357	SILD-----FSSKLDIVYSSDKFYKQ--TVNQLQPYSLVTLTNSDL-----KY 397	46;	Mismatches 135;	Indels 172;	Gaps 20;		
QY	156	ELLDDGPGT---QPVQFAIIG-----QPYVHKWTCSETV-----DTFCAY-----193						
DB	398	NALDLTNNKCLRLEPKLHVAGNLKGAYONNEIKHIVAISSAALSASYKADTVAKQGVGE 457						
QY	194	-VHSCFVD-----DGNGDIYVELINA-----212						
DB	458	FSHRLNTDIAGLASAIDMSTNTNSDLHFSNVFRSVAAPFTMTIDAHTNGKLTALWGEH 517						
QY	213	DCGALDKYLLNNLEYPPTDLMAQGEAHVKKYADRSOLFQCOISITTIKEPNSECVRPCQCE 272						
DB	518	TGOLYKFPILK-----AEPALFTFSDHKYSTSHHLVSRKSSISALEHKYSALLTP--AE 570						
QY	273	PGGFGAVKGT-----GAAAKPAAAOGLRLKK-----RSAEPENI 307						
DB	571	QGTGWLKTKQFNNEXSQDLDAVNTKDKIGVELTGRTLADLTLLDSPIKVPLLSEPINI 630						
QY	308	IDVRDINTLEISD 321						
DB	631	ID-----ALEMRD 638						
RESULT 21								
US-09-870-759-128								
Sequence 128, Application US/09870759								
Patent No. US2002017751A1								
GENERAL INFORMATION:								
APPLICANT: TERMAN, David S								
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT OF NEOPLASTIC DISEASE								
FILE REFERENCE: 870759								
CURRENT APPLICATION NUMBER: US/09/870,759								
CURRENT FILING DATE: 2002-01-14								
PRIOR APPLICATION NUMBER: US 60/208,128								
PRIOR FILING DATE: 2000-05-30								
NUMBER OF SEQ ID NOS: 166								
SOFTWARE: PatentIn version 3.1								
SEQ ID NO 128								
LENGTH: 4563								
TYPE: PRT								
ORGANISM: Homo sapiens								
US-09-870-759-128								
Query Match								
Best Local Similarity 18.7%;								
Matches 81; Conservative 46; Mismatches 135; Indels 172; Gaps 20;								
QY	9	CTTLLALSTIPVDNVEGEPEIECGPTSTITINFNTAFEGH-----51						
DB	1662	CSILV-----LENELNALDELGISGASMKLTITNGRFRREHNAKESLDGKAALTELSTL 1710						
QY	52	-----VYVKGLYD-----OEGCR--NDEGGQVQAG:SLPRDSCNVARTSLNPRGI 95						
DB	1711	GSAYQAMILGVDSKNIFNFVKVSOEGKLKLSNDMMGSA---EMKFD-----HTNSLNTIAGL 1762						
QY	96	FTTTTAVISFHPLEFVKRVKADRAVROCFYMBADKTVSAQIEVSEFTTAFOQIIVMPVCRY 155						

```
Db 1924 TGLIYSFLK-----AEPALFTSHDYKSTSHLVSRKISALAEHKVSAITLP--AE 1976
QY 273 POGFGAVKTG-----GAAKPAAAOURLKK-----RSEPEPNI 307
Db 1977 QTGWKTKTGFNNNYSODDAVNTKDKIGVELGRLADLTLLDSPIKVPLLSPEINI 2036
QY 308 IDVNTDINTLEISD 321
Db 2037 ID-----ALEMRD 2044

RESULT 22
US-09-802-640-32
; Sequence 32, Application US/09802640
; Publication No. US20030036057A1
; GENERAL INFORMATION:
; APPLICANT: Braun, Andreas
; APPLICANT: Bonsel Aruna
; APPLICANT: Kleyn Patrick
; TITLE OF INVENTION: GENES AND POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISEASE AND THEIR USE
; FILE REFERENCE: 24736-2048
; CURRENT APPLICATION NUMBER: US/09/802,640
; CURRENT FILING DATE: 2001-03-09
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 32
; LENGTH: 4563
; TYPE: PRF
; ORGANISM: Homo saplen
US-09-802-640-32

Query Match
Best Local Similarity 18.7%; Score 77; DB 9; Length 4563;
Matches 81; Conservative 46; Mismatches 135; Indels 172; Gaps 20;

QY 9 CTTLLAISTIPVDNGVEGEPEIECGPTSTINFNTRNAFEHG----- 51
Db 1662 CSDLV-----LENELNAELIGSGASMKLTNGRFRHNAKPSLDGKAALTEL 1710
QY 52 -----VYVKGLVD-----OEGGR--NDEGGROVAGSLPFDSCNVARTRSLNPRGI 95
Db 1711 GSAYQAMILGVDKSNINFEVSOEGLKLSNDMGSTA--EMKFD-----HTNSLNTAGL 1762
QY 96 FVTTTIVISFHPFVTKVDNARVOCFYMEADKTVSAQISEITTAFOQIYPMPEVCRY 155
Db 1763 SLD-----FSSKLDNTISSDKFYKQ--TVNLQQLPYSLVYTLNIDL-----KY 1803
QY 156 EIIDGCPFG---QPVQFALIG-----QPVYHKWTCDSFTV-----DFECAY--- 193
Db 1804 NALDLTNNNGRLRLEPLKLHVGAMIKGAYONNEIKHIAISSAALSASYKADTVAKQGYE 1863
QY 194 -VHSCFVD-----DGNDRIVEILNA----- 212
Db 1864 FSRRLNTDLAGLASAIDMSINYSDSLHSNVRSWAPPTMTIDAHNTNGKLAIMEH 1923
QY 213 DGCALDKYLLNLEPFTDLMAGQEAHVYKYADRSQLFYOCQISITIKRPNSECVRQCS 272
Db 1924 TGLYLSKFLTK-----AEPALFTSHDYKSTSHLVSRKISALAEHKVSAITLP--AE 1976
QY 273 POGFGAVKTG-----GAAKPAAAOURLKK-----RSEPEPNI 307
Db 1977 QTGWKTKTGFNNNYSODDAVNTKDKIGVELGRLADLTLLDSPIKVPLLSPEINI 2036
QY 308 IDVNTDINTLEISD 321
Db 2037 ID-----ALEMRD 2044

RESULT 23
US-09-738-626-4782
; Sequence 4782, Application US/09738626
; Publication No. US20020197605A1
```

```
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 4782
; LENGTH: 295
; TYPE: PRF
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-4782

Query Match
Best Local Similarity 20.8%; Score 76.5; DB 9; Length 295;
Matches 38; Conservative 26; Mismatches 72; Indels 47; Gaps 7;

QY 79 FDSCNNARTRSLNPRGIFVTTIVISFHP-----FVTKVDRAV--RQ 120
Db 22 FDSCE-----NPRKLWKSSTLVHLDHFIRDVRLTEERWNAIDFLTRKVGHTDDK 74
QY 121 CFYMEADKTVSAQIESEITTAFOQIYPMPCRYEILDGP---TGQPVQFALIGQPVY 177
Db 75 EFVLTD--TLGASMQTIVANNENAEATEDATEVGPFDVADAPLVQGGDIAFGAVGQPMW 133
QY 178 HKWTCSEVYDFCAVHSCFVDGNGDVEIILNADGALDKYLLNLEPFTDLMAGQEA 237
Db 134 VEGT-----VKDTGPNIPNARIIEWMECEDGELTVQYADERSAGR-A 175
QY 238 HUY 240
Db 176 HLY 178

RESULT 24
US-09-992-598-124
; Sequence 124, Application US/0992598
; Patent No. US20020160384A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi J.
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Kijavita, Ivar J.
; APPLICANT: Napier, Mary A.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
```

APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K.
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: P2730PIC20
CURRENT APPLICATION NUMBER: US/09/992,598
CURRENT FILING DATE: 2001-11-14
PRIOR APPLICATION NUMBER: 60/049787
PRIOR FILING DATE: 1997-06-16
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/065186
PRIOR FILING DATE: 1997-11-12
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066770
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/075945
PRIOR FILING DATE: 1998-02-25
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/084600
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/087106
PRIOR FILING DATE: 1998-05-28
PRIOR APPLICATION NUMBER: 60/087607
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087609
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087759
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087827
PRIOR FILING DATE: 1998-06-03
PRIOR APPLICATION NUMBER: 60/088021
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088025
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088026
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088028
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088029
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088030
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088033
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088326
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088167
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088202
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088212
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088217
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088655
PRIOR FILING DATE: 1998-06-09
PRIOR APPLICATION NUMBER: 60/088734
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088738
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088742
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088810
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088824
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088826
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088858
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088861
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088876
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/089105
PRIOR FILING DATE: 1998-06-12
PRIOR APPLICATION NUMBER: 60/089440
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089512
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089514
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089532
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089538
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089598
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089599
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089600
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089653
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089801
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089907
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089908
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089947
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/089948
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/089952
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/090246
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090252
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090254
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090349
PRIOR FILING DATE: 1998-06-23
PRIOR APPLICATION NUMBER: 60/090355
PRIOR FILING DATE: 1998-06-23
PRIOR APPLICATION NUMBER: 60/090429
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090431
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090435
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090444
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090445
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090472
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090535
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090540
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090542
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090557
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090676

```

; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090678
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090690
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090694
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090695
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090696
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090862
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/090863
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/091360
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091478
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091544
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091519
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091626
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091633
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091978
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/091982
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/092182
; PRIOR FILING DATE: 1998-07-09

```

```

Query Match          3.8%; Score 76.5; DB 9; Length 1184;
Best Local Similarity 21.2%; Pred. No. 77;
Matches 72; Conservative 47; Mismatches 123; Indels 97; Gaps 19;

```

```

QY 17 YSLPVDNGEGERETCGTSTTINFNTNAEGHYV-----KSLYDQEGGRNDE 67
DB 660 FSVDFDEVTSEB-LNAGVYKHL-D-STQVKMEHISTVKLSLNDTGLMEEGFKRE 717
QY 68 GGRQVAGISLPFDCSVARTSLNPGITVTTTIVSIFPLVTKYD-----RAYR 118
DB 718 NGR-----NKRDRF-----FLVGNLEIRERRLNLDYVESRRCFVKVRAVR 760
QY 119 VOCFTWEDKTVSAQIE--VSEITTAFTQIVPMPVC--RYEILDGSGTGPVQFAIG 173
DB 761 SERF-----LPSEQIOGVIVSIVNLEPRTGFLSNPRAMGRDSTVTGFGNACV----- 808
QY 174 QPVYHKMTCDSEVTDFCAVYHSCFVDGNGDPTVEIL-----NADGALDKYLLNLEY 227
DB 809 -PAF-----CDQOSPDAVSAYVLASLA---GELQAVESSPKFNMAIGVPPYLNKNTLY 859
QY 228 -PTDLMAQEAHVYKVAARSLQFYQCOISITIKENSECVRPQCEPQG--FGAVKTCGA 284
DB 860 RTD--HEDRVKTKA-----FQTSMAKPRNS-----AESNGPIYAFENLRAC 902
QY 285 AAKPAAQAQLRLKRSAPENIIVRTDITLTLEISDN 323
DB 903 EEAAPSAHFRFYQ-----ISGDRDNTVTFPNEDD 933

```

```

RESULT 25
US-09-989-293A-124
; Sequence 124, Application US/09989293A
; Patent No. US20020177164A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi J.
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnovers, Luc
; APPLICANT: Baton, Dan L.

```

```

; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Kijavini, Ivar J.
; APPLICANT: Napier, Mary A.
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: P2730PIC66
; CURRENT APPLICATION NUMBER: US/09/989,293A
; PRIOR FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/049787
; PRIOR FILING DATE: 1997-06-16
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/065186
; PRIOR FILING DATE: 1997-11-12
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/067770
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/075945
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/084600
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/087106
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087609
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087759
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087827
; PRIOR FILING DATE: 1998-06-03
; PRIOR APPLICATION NUMBER: 60/088021
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088025
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088026
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088028
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088029
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088030
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088033
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088326
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088167
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088202
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088212

```


Db 903 EAPPSAHPHYO-----ICDRXNTVTPNEDD 933

RESULT 26
US-09-989-735-124
Sequence 124, Application US/09989735
Publication No. US20020193299A1

GENERAL INFORMATION:
APPLICANT: Ashkenazi, Avi J.
APPLICANT: Baker, Kevin P.
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Epton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Fong, Sherman
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerltsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Grimaldi, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Kljavin, Ivar J.
APPLICANT: Napier, Mary A.
APPLICANT: Pan, James
APPLICANT: Raoul, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K.
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin

TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: P2730PIC61
CURRENT APPLICATION NUMBER: US/09/989,735
CURRENT FILING DATE: 2001-11-19
PRIOR APPLICATION NUMBER: 60/049787
PRIOR FILING DATE: 1997-06-16
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/065186
PRIOR FILING DATE: 1997-11-12
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066770
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/075945
PRIOR FILING DATE: 1998-02-25
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/084600
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/087106
PRIOR FILING DATE: 1998-05-28
PRIOR APPLICATION NUMBER: 60/087607
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087609
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087759
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087827
PRIOR FILING DATE: 1998-06-03
PRIOR APPLICATION NUMBER: 60/088021
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088025
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088026
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088028
PRIOR FILING DATE: 1998-06-04

PRIOR APPLICATION NUMBER: 60/088029
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088030
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088033
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088326
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088167
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088202
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088212
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088217
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088655
PRIOR FILING DATE: 1998-06-09
PRIOR APPLICATION NUMBER: 60/088734
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088738
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088742
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088810
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088824
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088826
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088858
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088861
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088876
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/089105
PRIOR FILING DATE: 1998-06-12
PRIOR APPLICATION NUMBER: 60/089440
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089512
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089514
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089532
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089538
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089598
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089599
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089600
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089653
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089801
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089907
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089908
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089947
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/089948
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/089952
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/090246
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090252
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090254

```

: PRIOR FILING DATE: 1998-06-22
: PRIOR APPLICATION NUMBER: 60/090349
: PRIOR FILING DATE: 1998-06-23
: PRIOR APPLICATION NUMBER: 60/090355
: PRIOR FILING DATE: 1998-06-23
: PRIOR APPLICATION NUMBER: 60/090429
: PRIOR FILING DATE: 1998-06-24
: PRIOR APPLICATION NUMBER: 60/090431
: PRIOR FILING DATE: 1998-06-24
: PRIOR APPLICATION NUMBER: 60/090435
: PRIOR FILING DATE: 1998-06-24
: PRIOR APPLICATION NUMBER: 60/090444
: PRIOR FILING DATE: 1998-06-24
: PRIOR APPLICATION NUMBER: 60/090445
: PRIOR FILING DATE: 1998-06-24
: PRIOR APPLICATION NUMBER: 60/090472
: PRIOR FILING DATE: 1998-06-24
: PRIOR APPLICATION NUMBER: 60/090535
: PRIOR FILING DATE: 1998-06-24
: PRIOR APPLICATION NUMBER: 60/090540
: PRIOR FILING DATE: 1998-06-24
: PRIOR APPLICATION NUMBER: 60/090542
: PRIOR FILING DATE: 1998-06-24
: PRIOR APPLICATION NUMBER: 60/090557
: PRIOR FILING DATE: 1998-06-24
: PRIOR APPLICATION NUMBER: 60/090676
: PRIOR FILING DATE: 1998-06-25
: PRIOR APPLICATION NUMBER: 60/090678
: PRIOR FILING DATE: 1998-06-25
: PRIOR APPLICATION NUMBER: 60/090690
: PRIOR FILING DATE: 1998-06-25
: PRIOR APPLICATION NUMBER: 60/090694
: PRIOR FILING DATE: 1998-06-25
: PRIOR APPLICATION NUMBER: 60/090695
: PRIOR FILING DATE: 1998-06-25
: PRIOR APPLICATION NUMBER: 60/090696
: PRIOR FILING DATE: 1998-06-25
: PRIOR APPLICATION NUMBER: 60/090862
: PRIOR FILING DATE: 1998-06-26
: PRIOR APPLICATION NUMBER: 60/090863
: PRIOR FILING DATE: 1998-06-26
: PRIOR APPLICATION NUMBER: 60/091360
: PRIOR FILING DATE: 1998-07-01
: PRIOR APPLICATION NUMBER: 60/091478
: PRIOR FILING DATE: 1998-07-02
: PRIOR APPLICATION NUMBER: 60/091544
: PRIOR FILING DATE: 1998-07-01
: PRIOR APPLICATION NUMBER: 60/091519
: PRIOR FILING DATE: 1998-07-02
: PRIOR APPLICATION NUMBER: 60/091626
: PRIOR FILING DATE: 1998-07-02
: PRIOR APPLICATION NUMBER: 60/091633
: PRIOR FILING DATE: 1998-07-02
: PRIOR APPLICATION NUMBER: 60/091978
: PRIOR FILING DATE: 1998-07-07
: PRIOR APPLICATION NUMBER: 60/091982
: PRIOR FILING DATE: 1998-07-07
: PRIOR APPLICATION NUMBER: 60/092182
: PRIOR FILING DATE: 1998-07-09

Query Match          3.8%; Score 76.5; DB 9; Length 1184;
Best Local Similarity 21.2%; Pred. No. 77;
Matches 72; Conservative 47; Mismatches 123; Indels 97; Gaps 19;

QY 17 YSPVNGVGEPEIECGPISITINENTRNAFEGHYV-----KGLYDQSGCRNDE 67
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 660 FSVDFREVTSEF-1NAGKAKVHD-STOVKMPHISTVKLMSLNPDTGLMEEGDPKFE 717
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 68 GGRVAGISLPFSCNATRSINPRGIFVTTTVVISFHFPLEFYTKVD-----RAYR 118
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 718 NQRR-----NKRERT-----FLVGNLEIRERLFJNDVDSRRRCFYKVRAYR 760
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 119 VOCFYMEADKTVSAQIE--VSEITTAFOFOIYVMPVC--RYEILDGPGTGOVQFAIIG 173
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```

DB 761 SERF-----LPSEQIQGVNIVSINLEPRTFLSPRAMGRFDSYITPGNACV----- 808
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 174 QPVYHKWTCDSEYVDTFCAVYHSCFVDDGNDTYEIL-----NADGALDKRYLLNNLEY 227
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 809 -PAF-----CDQSPDPAVSAYVLAIA-----GEELQAVESSPKFPNNAIGVPOPLYINKLNY 859
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 228 -PTDLMAGQEAHYKYNADRSOLFVOCQISITIKKEPNSFCVAPQCSPEQG--FGAVKTTGA 284
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 860 RRTD---HEDPRVKTA-----FOISMKPPRNS-----AESNGPIYAFENLRAC 902
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 285 AAKPAAQAOLRLKKRSAPENIIDVPTDITLLEISDN 323
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 903 EAEPSSAHFRFYQ-----IESDRDYNTVTPENEDD 933
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 27
US-09-990-444-124
: Sequence 124, Application US/09990444
: Publication No. US20020193300A1
: GENERAL INFORMATION:
: APPLICANT: Ashkenazi, Avi J.
: APPLICANT: Baker, Kevin P.
: APPLICANT: Botstein, David
: APPLICANT: Desnoyers, Luc
: APPLICANT: Eaton, Dan L.
: APPLICANT: Ferrara, Napoleone
: APPLICANT: Fong, Sherman
: APPLICANT: Gerber, Hanspeter
: APPLICANT: Gerritsen, Mary E.
: APPLICANT: Goddard, Audrey
: APPLICANT: Godowski, Paul J.
: APPLICANT: Grimaldi, J. Christopher
: APPLICANT: Gurney, Austin L.
: APPLICANT: Kijavlin, Ivar J.
: APPLICANT: Kijavlin, Ivar J.
: APPLICANT: Napier, Mary A.
: APPLICANT: Pan, James
: APPLICANT: Paoni, Nicholas F.
: APPLICANT: Roy, Margaret Ann
: APPLICANT: Stewart, Timothy A.
: APPLICANT: Tumas, Daniel
: APPLICANT: Watanabe, Colin K.
: APPLICANT: Williams, P. Mickey
: APPLICANT: Wood, William I.
: APPLICANT: Zhang, Zemin
: TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
: FILE REFERENCE: P2730P1C19
: CURRENT APPLICATION NUMBER: US/09/990,444
: CURRENT FILING DATE: 2001-11-14
: PRIOR APPLICATION NUMBER: 60/049787
: PRIOR FILING DATE: 1997-06-16
: PRIOR APPLICATION NUMBER: 60/062250
: PRIOR FILING DATE: 1997-10-17
: PRIOR APPLICATION NUMBER: 60/065186
: PRIOR FILING DATE: 1997-11-12
: PRIOR APPLICATION NUMBER: 60/065311
: PRIOR FILING DATE: 1997-11-13
: PRIOR APPLICATION NUMBER: 60/066770
: PRIOR FILING DATE: 1997-11-24
: PRIOR APPLICATION NUMBER: 60/075945
: PRIOR FILING DATE: 1998-02-25
: PRIOR APPLICATION NUMBER: 60/078910
: PRIOR FILING DATE: 1998-03-20
: PRIOR APPLICATION NUMBER: 60/083322
: PRIOR FILING DATE: 1998-04-28
: PRIOR APPLICATION NUMBER: 60/084600
: PRIOR FILING DATE: 1998-05-07
: PRIOR APPLICATION NUMBER: 60/087106
: PRIOR FILING DATE: 1998-05-28
: PRIOR APPLICATION NUMBER: 60/087607
: PRIOR FILING DATE: 1998-06-02
: PRIOR APPLICATION NUMBER: 60/087609
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```


;; PRIOR FILING DATE: 1998-06-17
;; PRIOR APPLICATION NUMBER: 60/089598
;; PRIOR FILING DATE: 1998-06-17
;; PRIOR APPLICATION NUMBER: 60/089599
;; PRIOR FILING DATE: 1998-06-17
;; PRIOR APPLICATION NUMBER: 60/089600
;; PRIOR FILING DATE: 1998-06-17
;; PRIOR APPLICATION NUMBER: 60/089653
;; PRIOR FILING DATE: 1998-06-17
;; PRIOR APPLICATION NUMBER: 60/089801
;; PRIOR FILING DATE: 1998-06-18
;; PRIOR APPLICATION NUMBER: 60/089907
;; PRIOR FILING DATE: 1998-06-18
;; PRIOR APPLICATION NUMBER: 60/089908
;; PRIOR FILING DATE: 1998-06-18
;; PRIOR APPLICATION NUMBER: 60/089947
;; PRIOR FILING DATE: 1998-06-19
;; PRIOR APPLICATION NUMBER: 60/089948
;; PRIOR FILING DATE: 1998-06-19
;; PRIOR APPLICATION NUMBER: 60/089952
;; PRIOR FILING DATE: 1998-06-19
;; PRIOR APPLICATION NUMBER: 60/090246
;; PRIOR FILING DATE: 1998-06-22
;; PRIOR APPLICATION NUMBER: 60/090252
;; PRIOR FILING DATE: 1998-06-22
;; PRIOR APPLICATION NUMBER: 60/090254
;; PRIOR FILING DATE: 1998-06-22
;; PRIOR APPLICATION NUMBER: 60/090349
;; PRIOR FILING DATE: 1998-06-23
;; PRIOR APPLICATION NUMBER: 60/090355
;; PRIOR FILING DATE: 1998-06-23
;; PRIOR APPLICATION NUMBER: 60/090429
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090431
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090435
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090444
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090445
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090472
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090535
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090540
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090542
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090557
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090676
;; PRIOR FILING DATE: 1998-06-25
;; PRIOR APPLICATION NUMBER: 60/090678
;; PRIOR FILING DATE: 1998-06-25
;; PRIOR APPLICATION NUMBER: 60/090690
;; PRIOR FILING DATE: 1998-06-25
;; PRIOR APPLICATION NUMBER: 60/090694
;; PRIOR FILING DATE: 1998-06-25
;; PRIOR APPLICATION NUMBER: 60/090695
;; PRIOR FILING DATE: 1998-06-25
;; PRIOR APPLICATION NUMBER: 60/090696
;; PRIOR FILING DATE: 1998-06-25
;; PRIOR APPLICATION NUMBER: 60/090862
;; PRIOR FILING DATE: 1998-06-26
;; PRIOR APPLICATION NUMBER: 60/090863
;; PRIOR FILING DATE: 1998-06-26
;; PRIOR APPLICATION NUMBER: 60/091360
;; PRIOR FILING DATE: 1998-07-01
;; PRIOR APPLICATION NUMBER: 60/091478
;; PRIOR FILING DATE: 1998-07-02
;; PRIOR APPLICATION NUMBER: 60/091544
;; PRIOR FILING DATE: 1998-07-01

;; PRIOR APPLICATION NUMBER: 60/091519
;; PRIOR FILING DATE: 1998-07-02
;; PRIOR APPLICATION NUMBER: 60/091626
;; PRIOR FILING DATE: 1998-07-02
;; PRIOR APPLICATION NUMBER: 60/091633
;; PRIOR FILING DATE: 1998-07-02
;; PRIOR APPLICATION NUMBER: 60/091978
;; PRIOR FILING DATE: 1998-07-07
;; PRIOR APPLICATION NUMBER: 60/091982
;; PRIOR FILING DATE: 1998-07-07
;; PRIOR APPLICATION NUMBER: 60/092182
;; PRIOR FILING DATE: 1998-07-09

Query Match 3.8%; Score 76.5; DB 9; Length 1184;
Best Local Similarity 21.2%; Pred. No. 77;
Matches 72; Conservative 47; Mismatches 123; Indels 97; Gaps 19;

QY 17 YSTIVDNGVEGEPEIEGPTISITINETRNAFEHVV-----KGLYDEGCRNDE 67
DB 660 FSVDFRDEVSEPLNAGKVKVHLD-STOVKMPHISTVKLMSLNPDTGLMEEGDKFFE 717
QY 68 GGRQVAGISLPDSCNVAFTRSLSNPGIFVTTVVISFHPLEVTXVD-----RAYR 118
DB 718 NQRR-----AKREDRT-----FLVGNLEIERKRLFNLDVPESSRCFYKVRAYR 760
QY 119 VQCFYMEADKTVSAQIE---VSEITTAFOQIYMPVC--RYEILDGPGTGPQYFAIIG 173
DB 761 SERP-----LPSEQIQGVIVISVINEPRTGFLSNPRAMGRFDSVITGPGACV----- 808
QY 174 QPYVHKWTCSEYVDIFCAVYHSCFVDGDCGVDELL-----NADGCALDKTLLNNLEY 227
DB 809 -PAF---CDQSPDASAVVLASLA---GEELOAVESSPKRPNNAIGVPOPYLKNLNY 859
QY 228 -PTDLMAGQEAHYKXKADRSQLFYOCOISITIKEPNSCEVRCPCSEPG--FGAVVTGGA 284
DB 860 RRTD---HEPRVKKTA-----FOISAKRPPS-----AESNGPTTAFENLRAC 902
QY 285 AAKPAAAOQLRLKKRSAPENIIDVPTDINTLEISDN 323
DB 903 EEAPPSAAHFRFYQ-----IEGDRYDYNVTVPFNMEDD 933

RESULT 29
US-09-990-436-124
; Sequence 124, Application US/09990436
; Publication No. US20020198148A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi J.
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Napier, Mary A.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same

[illegible]

```

; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090696
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090862
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/090863
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/091360
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091478
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091544
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091519
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091626
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091633
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091978
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/091982
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/092182
; PRIOR FILING DATE: 1998-07-09

```

Query Match 3.8%; Score 76.5; DB 9; Length 1184;
Best Local Similarity 21.2%; Pred. No. 77;

Matches 72; Conservative 47; Mismatches 123; Indels 97; Gaps 19;

```

QY 17 YSIPVNGVGEPEICGTSITINFTNNAFGHYV-----KGIYDEGRNDE 67
DB 660 FSYDFDEVTSSE-LNAGVYKHALD-STGVKMEHISTYKLSLNDPTGLMEBGDFKE 717
QY 68 GGRVYAGISLPDSCNVAATRSINPRGIFVTTTVISFHLFTYKVD-----RAYR 118
DB 718 NQNR-----NKREDRT-----FLVGNLEIRERRLFLUDVPESRRCVYKRAIR 760
QY 119 VOCFWEADKYVSAQIE--VSEITTAFOYQIVPMFVC--RREIIDGFTGQVPQPAIG 173
DB 761 SERF-----LPSEQIOGVIVISINLEPRTGFLSNPRAMGRFDSYITGNACV----- 808
QY 174 QPVYHKMCDSEVDVDFCAVHSCFVDDGNGDVEL-----NADGCLDKYLLNLEY 227
DB 809 -DAF-----CDDSDPAVYSVYVLAISLA-----GEELQAVESSPFENPAIGVPOYLNKLMY 859
QY 228 -PTDLMAOEAHVYKADASQLFYQCQISITIKENPSECVRPQCSPPQG--FGAVYKGA 284
DB 860 RRTD---HEDPRVYKTA-----FQISMAKPRPNS-----AESNGPIYAFENLRAC 902
QY 285 AAKPAAAOQLRLKKRSAPENIIDVRTDINTLEISDDN 323
DB 903 EEAPPSAAHFRFYQ-----IEGDRYDYNTVNFENEDD 933

```

RESULT 30
US-09-991-181-124
; Sequence 124, Application US/09991181
; Publication No. US20020197615A1

GENERAL INFORMATION:

```

; APPLICANT: Ashkenazi, Avi J.
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.

```

```

; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Napier, Mary A.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P27301C53
; CURRENT APPLICATION NUMBER: US/09/991,181
; PRIOR FILING DATE: 2001-11-16
; PRIOR APPLICATION NUMBER: 60/049787
; PRIOR FILING DATE: 1997-06-16
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/065186
; PRIOR FILING DATE: 1997-11-12
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/067770
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/075945
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/084600
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/087106
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087609
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087759
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087827
; PRIOR FILING DATE: 1998-06-03
; PRIOR APPLICATION NUMBER: 60/088021
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088025
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088026
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088028
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088029
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088030
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088033
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088326
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088167
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088202
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088212
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088217
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088655
; PRIOR FILING DATE: 1998-06-09
; PRIOR APPLICATION NUMBER: 60/088734
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088738

```



```

PRIORITY APPLICATION NUMBER: 60/090542
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090557
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090676
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090678
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090690
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090694
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090695
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090696
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090862
PRIOR FILING DATE: 1998-06-26
PRIOR APPLICATION NUMBER: 60/090863
PRIOR FILING DATE: 1998-06-26
PRIOR APPLICATION NUMBER: 60/091360
PRIOR FILING DATE: 1998-07-01
PRIOR APPLICATION NUMBER: 60/091478
PRIOR FILING DATE: 1998-07-02
PRIOR APPLICATION NUMBER: 60/091544
PRIOR FILING DATE: 1998-07-01
PRIOR APPLICATION NUMBER: 60/091519
PRIOR FILING DATE: 1998-07-02
PRIOR APPLICATION NUMBER: 60/091626
PRIOR FILING DATE: 1998-07-02
PRIOR APPLICATION NUMBER: 60/091633
PRIOR FILING DATE: 1998-07-02
PRIOR APPLICATION NUMBER: 60/091978
PRIOR FILING DATE: 1998-07-07
PRIOR APPLICATION NUMBER: 60/091982
PRIOR FILING DATE: 1998-07-07
PRIOR APPLICATION NUMBER: 60/092182
PRIOR FILING DATE: 1998-07-09

Query Match          3.8%; Score 76.5; DB 9; Length 1184;
Best Local Similarity 21.2%; Pred. No. 77;
Matches 72; Conservative 47; Mismatches 123; Indels 97; Gaps 19.

QY      17 YSIPDNGVESEPEIECCPTSTITINFNTRNAFECHVY-----KGLYDEGGRNDE 67
       ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      660 FSVDPRDEVTEBP-LNAKKVKVHLD-STQVKMKREHISTYKMLSLNDPIGLMWEEDGFKE 717

QY      68 GGROVAGISLPDDSCNVARTNSLNRGIFVTTVVISFHPLEFVTYVD-----RAYR 118
       |:::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      718 NQR-----NKREDRT-----FLVGNLEIRERRLNLDPESRRCFVKRAVR 760

QY      119 VQCEFMADKYVSAQIE---VSEITTAFOYQIVMPVC--RREILDGGTGGPOQFAIG 173
       |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      761 SER-----LPSEIDIGGVISVINLEPRBTGLSNPRAMGRPDVSITGNGACV----- 808

QY      174 QPVYHKWTCDSEVDTFCFAVHSCFVDGNGDFTVEL-----NADGCALDKYLINLTLEY 227
       |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      809 -PAF-----CDOOSPAYIASYVLASLA-----GELOAVESSPKFNNAIGAPPYLNKLNY 859

QY      228 -PTDMAQOEAHVYKYADRSQLFYQQGISITIKEPNSECVQCSEPGG--FGAVKTGA 284
       ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      860 RRTD---HEDRPVKKA-----FQLSMKKPRPN-----AEESNGPIYAFENLRAC 902

QY      285 AAKPAAAOQLRLKKRSAPENIIDVRINDINTLEISDN 323
       |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      903 EAPPSAHRFRYO-----TGDRIYDNIPTPNEDD 933

RESULT 31
US-09-993-687-124
Sequence 124, Application US/09993687
Publication No. US20020198149A1
GENERAL INFORMATION:
```

APPLICANT: Ashkenazi, Avi J.
 APPLICANT: Baker, Kevin P.
 APPLICANT: Botstein, David
 APPLICANT: Desnoyers, Luc
 APPLICANT: Eaton, Dan L.
 APPLICANT: Ferrara, Napoleone
 APPLICANT: Fong, Sherman
 APPLICANT: Gerber, Hanspeter
 APPLICANT: Gerltsen, Mary E.
 APPLICANT: Goddard, Audrey
 APPLICANT: Godowski, Paul J.
 APPLICANT: Grimaldi, J. Christopher
 APPLICANT: Gurney, Austin L.
 APPLICANT: Kijavyn, Ivar J.
 APPLICANT: Napier, Mary A.
 APPLICANT: Pan, James
 APPLICANT: Paoni, Nicholas F.
 APPLICANT: Roy, Margaret Ann
 APPLICANT: Stewart, Timothy A.
 APPLICANT: Tumas, Daniel
 APPLICANT: Watanabe, Colin K.
 APPLICANT: Williams, P. Mickey
 APPLICANT: Wood, William I.
 APPLICANT: Zhang, Zemin
 TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
 FILE OF INVENTION: Acids Encoding the Same
 FILE REFERENCE: P2730P1C11
 CURRENT APPLICATION NUMBER: US/09/993,687
 PRIOR FILING DATE: 2002-11-14
 PRIOR APPLICATION NUMBER: 60/049787
 PRIOR FILING DATE: 1997-06-16
 PRIOR APPLICATION NUMBER: 60/062250
 PRIOR FILING DATE: 1997-10-17
 PRIOR APPLICATION NUMBER: 60/065186
 PRIOR FILING DATE: 1997-11-12
 PRIOR APPLICATION NUMBER: 60/065311
 PRIOR FILING DATE: 1997-11-13
 PRIOR APPLICATION NUMBER: 60/066770
 PRIOR FILING DATE: 1997-11-24
 PRIOR APPLICATION NUMBER: 60/075945
 PRIOR FILING DATE: 1998-02-25
 PRIOR APPLICATION NUMBER: 60/078910
 PRIOR FILING DATE: 1998-03-20
 PRIOR APPLICATION NUMBER: 60/083322
 PRIOR FILING DATE: 1998-04-28
 PRIOR APPLICATION NUMBER: 60/084600
 PRIOR FILING DATE: 1998-05-07
 PRIOR APPLICATION NUMBER: 60/087106
 PRIOR FILING DATE: 1998-05-28
 PRIOR APPLICATION NUMBER: 60/087607
 PRIOR FILING DATE: 1998-06-02
 PRIOR APPLICATION NUMBER: 60/087609
 PRIOR FILING DATE: 1998-06-02
 PRIOR APPLICATION NUMBER: 60/087759
 PRIOR FILING DATE: 1998-06-02
 PRIOR APPLICATION NUMBER: 60/087827
 PRIOR FILING DATE: 1998-06-03
 PRIOR APPLICATION NUMBER: 60/088021
 PRIOR FILING DATE: 1998-06-04
 PRIOR APPLICATION NUMBER: 60/088025
 PRIOR FILING DATE: 1998-06-04
 PRIOR APPLICATION NUMBER: 60/088026
 PRIOR FILING DATE: 1998-06-04
 PRIOR APPLICATION NUMBER: 60/088028
 PRIOR FILING DATE: 1998-06-04
 PRIOR APPLICATION NUMBER: 60/088029
 PRIOR FILING DATE: 1998-06-04
 PRIOR APPLICATION NUMBER: 60/088030
 PRIOR FILING DATE: 1998-06-04
 PRIOR APPLICATION NUMBER: 60/088033
 PRIOR FILING DATE: 1998-06-04
 PRIOR APPLICATION NUMBER: 60/088326
 PRIOR FILING DATE: 1998-06-04

PRIOR APPLICATION NUMBER: 60/088167
 PRIOR FILING DATE: 1998-06-05
 PRIOR APPLICATION NUMBER: 60/088202
 PRIOR FILING DATE: 1998-06-05
 PRIOR APPLICATION NUMBER: 60/088212
 PRIOR FILING DATE: 1998-06-05
 PRIOR APPLICATION NUMBER: 60/088217
 PRIOR FILING DATE: 1998-06-05
 PRIOR APPLICATION NUMBER: 60/088655
 PRIOR FILING DATE: 1998-06-09
 PRIOR APPLICATION NUMBER: 60/088734
 PRIOR FILING DATE: 1998-06-10
 PRIOR APPLICATION NUMBER: 60/088738
 PRIOR FILING DATE: 1998-06-10
 PRIOR APPLICATION NUMBER: 60/088742
 PRIOR FILING DATE: 1998-06-10
 PRIOR APPLICATION NUMBER: 60/088810
 PRIOR FILING DATE: 1998-06-10
 PRIOR APPLICATION NUMBER: 60/088824
 PRIOR FILING DATE: 1998-06-10
 PRIOR APPLICATION NUMBER: 60/088826
 PRIOR FILING DATE: 1998-06-10
 PRIOR APPLICATION NUMBER: 60/088858
 PRIOR FILING DATE: 1998-06-11
 PRIOR APPLICATION NUMBER: 60/088861
 PRIOR FILING DATE: 1998-06-11
 PRIOR APPLICATION NUMBER: 60/088876
 PRIOR FILING DATE: 1998-06-11
 PRIOR APPLICATION NUMBER: 60/089105
 PRIOR FILING DATE: 1998-06-12
 PRIOR APPLICATION NUMBER: 60/089440
 PRIOR FILING DATE: 1998-06-16
 PRIOR APPLICATION NUMBER: 60/089512
 PRIOR FILING DATE: 1998-06-16
 PRIOR APPLICATION NUMBER: 60/089514
 PRIOR FILING DATE: 1998-06-16
 PRIOR APPLICATION NUMBER: 60/089532
 PRIOR FILING DATE: 1998-06-17
 PRIOR APPLICATION NUMBER: 60/089538
 PRIOR FILING DATE: 1998-06-17
 PRIOR APPLICATION NUMBER: 60/089598
 PRIOR FILING DATE: 1998-06-17
 PRIOR APPLICATION NUMBER: 60/089599
 PRIOR FILING DATE: 1998-06-17
 PRIOR APPLICATION NUMBER: 60/089600
 PRIOR FILING DATE: 1998-06-17
 PRIOR APPLICATION NUMBER: 60/089653
 PRIOR FILING DATE: 1998-06-17
 PRIOR APPLICATION NUMBER: 60/089801
 PRIOR FILING DATE: 1998-06-18
 PRIOR APPLICATION NUMBER: 60/089907
 PRIOR FILING DATE: 1998-06-18
 PRIOR APPLICATION NUMBER: 60/089908
 PRIOR FILING DATE: 1998-06-18
 PRIOR APPLICATION NUMBER: 60/089947
 PRIOR FILING DATE: 1998-06-19
 PRIOR APPLICATION NUMBER: 60/089948
 PRIOR FILING DATE: 1998-06-19
 PRIOR APPLICATION NUMBER: 60/089952
 PRIOR FILING DATE: 1998-06-19
 PRIOR APPLICATION NUMBER: 60/090246
 PRIOR FILING DATE: 1998-06-22
 PRIOR APPLICATION NUMBER: 60/090252
 PRIOR FILING DATE: 1998-06-22
 PRIOR APPLICATION NUMBER: 60/090254
 PRIOR FILING DATE: 1998-06-22
 PRIOR APPLICATION NUMBER: 60/090349
 PRIOR FILING DATE: 1998-06-23
 PRIOR APPLICATION NUMBER: 60/090355
 PRIOR FILING DATE: 1998-06-23
 PRIOR APPLICATION NUMBER: 60/090429
 PRIOR FILING DATE: 1998-06-24
 PRIOR APPLICATION NUMBER: 60/090431

;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090435
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090444
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090445
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090472
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090535
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090540
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090542
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090557
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090676
;; PRIOR FILING DATE: 1998-06-25
;; PRIOR APPLICATION NUMBER: 60/090678
;; PRIOR FILING DATE: 1998-06-25
;; PRIOR APPLICATION NUMBER: 60/090690
;; PRIOR FILING DATE: 1998-06-25
;; PRIOR APPLICATION NUMBER: 60/090694
;; PRIOR FILING DATE: 1998-06-25
;; PRIOR APPLICATION NUMBER: 60/090695
;; PRIOR FILING DATE: 1998-06-25
;; PRIOR APPLICATION NUMBER: 60/090696
;; PRIOR FILING DATE: 1998-06-25
;; PRIOR APPLICATION NUMBER: 60/090862
;; PRIOR FILING DATE: 1998-06-26
;; PRIOR APPLICATION NUMBER: 60/090863
;; PRIOR FILING DATE: 1998-06-26
;; PRIOR APPLICATION NUMBER: 60/091360
;; PRIOR FILING DATE: 1998-07-01
;; PRIOR APPLICATION NUMBER: 60/091478
;; PRIOR FILING DATE: 1998-07-02
;; PRIOR APPLICATION NUMBER: 60/091544
;; PRIOR FILING DATE: 1998-07-01
;; PRIOR APPLICATION NUMBER: 60/091519
;; PRIOR FILING DATE: 1998-07-02
;; PRIOR APPLICATION NUMBER: 60/091626
;; PRIOR FILING DATE: 1998-07-02
;; PRIOR APPLICATION NUMBER: 60/091633
;; PRIOR FILING DATE: 1998-07-02
;; PRIOR APPLICATION NUMBER: 60/091978
;; PRIOR FILING DATE: 1998-07-07
;; PRIOR APPLICATION NUMBER: 60/091982
;; PRIOR FILING DATE: 1998-07-07
;; PRIOR APPLICATION NUMBER: 60/092182
;; PRIOR FILING DATE: 1998-07-09
Query Match 3.88; Score 76.5; DB 9; Length 1184;
Best Local Similarity 21.28; Pred. No. 77;
Matches 72; Conservative 47; Mismatches 123; Indels 97; Gaps 19;

Db 860 RRTD---HEPRVAKTA-----FOISMAKPPNS-----AESNGPIYAPENLRAC 902
QY 285 AAKPAAAOQLRLKRSAPENIIDVPTDINTLEISDN 323
Db 903 EAPPSAAHFRFYQ-----IEGDRXDVTYTFVFNED 933
RESULT 32
US-09-989-734-124
; Sequence 124, Application US/09989734
; Publication No. US20030003531A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi J.
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Falton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerlitsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Kijavlin, Ivar J.
; APPLICANT: Napier, Mary A.
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P230P1C64
; CURRENT FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: US/09/989,734
; PRIOR FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: 60/049787
; PRIOR FILING DATE: 1997-06-16
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/065186
; PRIOR FILING DATE: 1997-11-12
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066770
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/075945
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/084600
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/087106
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087609
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087759
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087827
; PRIOR FILING DATE: 1998-06-03
; PRIOR APPLICATION NUMBER: 60/088021
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088025

;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088026
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088028
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088029
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088030
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088033
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088326
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088167
;; PRIOR FILING DATE: 1998-06-05
;; PRIOR APPLICATION NUMBER: 60/088202
;; PRIOR FILING DATE: 1998-06-05
;; PRIOR APPLICATION NUMBER: 60/088212
;; PRIOR FILING DATE: 1998-06-05
;; PRIOR APPLICATION NUMBER: 60/088217
;; PRIOR FILING DATE: 1998-06-05
;; PRIOR APPLICATION NUMBER: 60/088655
;; PRIOR FILING DATE: 1998-06-09
;; PRIOR APPLICATION NUMBER: 60/088734
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088738
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088742
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088810
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088824
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088826
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088858
;; PRIOR FILING DATE: 1998-06-11
;; PRIOR APPLICATION NUMBER: 60/088861
;; PRIOR FILING DATE: 1998-06-11
;; PRIOR APPLICATION NUMBER: 60/088876
;; PRIOR FILING DATE: 1998-06-11
;; PRIOR APPLICATION NUMBER: 60/089105
;; PRIOR FILING DATE: 1998-06-12
;; PRIOR APPLICATION NUMBER: 60/089440
;; PRIOR FILING DATE: 1998-06-16
;; PRIOR APPLICATION NUMBER: 60/089512
;; PRIOR FILING DATE: 1998-06-16
;; PRIOR APPLICATION NUMBER: 60/089514
;; PRIOR FILING DATE: 1998-06-16
;; PRIOR APPLICATION NUMBER: 60/089532
;; PRIOR FILING DATE: 1998-06-17
;; PRIOR APPLICATION NUMBER: 60/089538
;; PRIOR FILING DATE: 1998-06-17
;; PRIOR APPLICATION NUMBER: 60/089598
;; PRIOR FILING DATE: 1998-06-17
;; PRIOR APPLICATION NUMBER: 60/089599
;; PRIOR FILING DATE: 1998-06-17
;; PRIOR APPLICATION NUMBER: 60/089600
;; PRIOR FILING DATE: 1998-06-17
;; PRIOR APPLICATION NUMBER: 60/089653
;; PRIOR FILING DATE: 1998-06-17
;; PRIOR APPLICATION NUMBER: 60/089801
;; PRIOR FILING DATE: 1998-06-18
;; PRIOR APPLICATION NUMBER: 60/089907
;; PRIOR FILING DATE: 1998-06-18
;; PRIOR APPLICATION NUMBER: 60/089908
;; PRIOR FILING DATE: 1998-06-18
;; PRIOR APPLICATION NUMBER: 60/089947
;; PRIOR FILING DATE: 1998-06-19
;; PRIOR APPLICATION NUMBER: 60/089948
;; PRIOR FILING DATE: 1998-06-19
;; PRIOR APPLICATION NUMBER: 60/089952
;; PRIOR FILING DATE: 1998-06-19

;; PRIOR APPLICATION NUMBER: 60/090246
;; PRIOR FILING DATE: 1998-06-22
;; PRIOR APPLICATION NUMBER: 60/090252
;; PRIOR FILING DATE: 1998-06-22
;; PRIOR APPLICATION NUMBER: 60/090254
;; PRIOR FILING DATE: 1998-06-22
;; PRIOR APPLICATION NUMBER: 60/090349
;; PRIOR FILING DATE: 1998-06-23
;; PRIOR APPLICATION NUMBER: 60/090355
;; PRIOR FILING DATE: 1998-06-23
;; PRIOR APPLICATION NUMBER: 60/090429
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090431
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090435
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090444
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090445
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090472
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090535
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090540
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090542
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090557
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090676
;; PRIOR FILING DATE: 1998-06-25
;; PRIOR APPLICATION NUMBER: 60/090678
;; PRIOR FILING DATE: 1998-06-25
;; PRIOR APPLICATION NUMBER: 60/090690
;; PRIOR FILING DATE: 1998-06-25
;; PRIOR APPLICATION NUMBER: 60/090694
;; PRIOR FILING DATE: 1998-06-25
;; PRIOR APPLICATION NUMBER: 60/090695
;; PRIOR FILING DATE: 1998-06-25
;; PRIOR APPLICATION NUMBER: 60/090696
;; PRIOR FILING DATE: 1998-06-25
;; PRIOR APPLICATION NUMBER: 60/090862
;; PRIOR FILING DATE: 1998-06-26
;; PRIOR APPLICATION NUMBER: 60/090863
;; PRIOR FILING DATE: 1998-06-26
;; PRIOR APPLICATION NUMBER: 60/091360
;; PRIOR FILING DATE: 1998-07-01
;; PRIOR APPLICATION NUMBER: 60/091478
;; PRIOR FILING DATE: 1998-07-02
;; PRIOR APPLICATION NUMBER: 60/091544
;; PRIOR FILING DATE: 1998-07-01
;; PRIOR APPLICATION NUMBER: 60/091519
;; PRIOR FILING DATE: 1998-07-02
;; PRIOR APPLICATION NUMBER: 60/091626
;; PRIOR FILING DATE: 1998-07-02
;; PRIOR APPLICATION NUMBER: 60/091633
;; PRIOR FILING DATE: 1998-07-02
;; PRIOR APPLICATION NUMBER: 60/091978
;; PRIOR FILING DATE: 1998-07-07
;; PRIOR APPLICATION NUMBER: 60/091982
;; PRIOR FILING DATE: 1998-07-07
;; PRIOR APPLICATION NUMBER: 60/092182
;; PRIOR FILING DATE: 1998-07-09

Query Match 3.8%, Score 76.5, DB 9, Length 1184;
Best Local Similarity 21.2%, Pred. No. 77;
Matches 72; Conservative 47; Mismatches 123; Indels 97; Gaps 19;

QY 17 YSIPDNVEGSEPLEDCPTSTITINFNRNAEEGHVY-----KGLVDQECGRNDE 67
DB 660 FSVDRDEVTSRP-LNACKVKVHLD-STOVKMPHISTIVKLMSLNPDRTGLWEEDGFKFE 717

Prior Application Number: 60/091519
Prior Filing Date: 1998-07-02
Prior Application Number: 60/091982
Prior Filing Date: 1998-07-07

Query Match 3.8%; Score 76.5; Db 9; Length 1184;
Best Local Similarity 21.2%; Pred.No.:77;
Matches 72; Conservative 47; Mismatches 123; Indels 97; Gaps 19

OY 17 VSIPVNDGVEGEPELEECGTSTITNFNTRNAEGHYV-----KGLYOQCGRANDE 67
Db :SVDPRDEVTSSEP-INAGKVKKHLD-STQVKMPREISTVKLTSLNDPLDGIAMEEEDFKFE 717
OY 68 GGROVAAGISLPPDSGNVARFRSLNRGRGFVTTVVISFHPRLFTVTYD-----RAAR 118
Db |: NORR-----NKREDRT-----FLVGNLERERRLLNLDVPESRCFFVKVRAIR 760
OY 119 VCCFEADKTYSQAIE---VSEITTAACTQTIVPAPVC--RYELIDGETGPVOFAIG 173
Db :SERF-----LPSEQIQGVLSVINLERPTGTLSPRAMGRDRSDVITGNACV----- 808
OY 174 QPVNYIKTCDESETVFPCAVALVHSCHVDGNGDTVEL-----NADGCALDKYLINNEY 227
Db :PAF----CDOQSPPAYSAUYLASLA-----GEILOAVESSPKFNMAICGPQVLINKLY 859
OY 228 -PTDMAGEAHVVYYUARSOLFYCOCISITKEPNSECVRQCSEPOG--FGAUKTGA 284
Db |||::|||::|||||:
OY 860 RTTD---HEDRVAKTA-----FOJSMKPFPNS-----ABESMGPIAFENLRMC 902
Db |:|||:|::|||
OY 285 AAKPAAAQLRLKLKRSAEPENIIDVRTDINTLETISDN 323
Db |||::|||:|::|||:
US-09-997-653-124
Sequence 124, Application US/09997653
Publication No. US20030008297A1
GENERAL INFORMATION:
APPPLICANT: Ashkenazi, Avi J.
APPPLICANT: Baker, Kevin P.
APPPLICANT: Botstein, David
APPPLICANT: Desnoyers, Luc
APPPLICANT: Eaton, Dan L.
APPPLICANT: Ferrara, Napoleone
APPPLICANT: Fong, Sherman
APPPLICANT: Gerber, Hanspeter
APPPLICANT: Gerritsen, Mary E.
APPPLICANT: Goddard, Audrey
APPPLICANT: Godowski, Paul J.
APPPLICANT: Grimaldi, J.Christopher
APPPLICANT: Gurney,Austin L.
APPPLICANT: Kijavini,Ivar J.
APPPLICANT: Napier,Mary A.
APPPLICANT: Pan,JAMES
APPPLICANT: Paoni,Nicholas F.
APPPLICANT: Roy, Margaret Ann
APPPLICANT: Stewart,Timothy A.
APPPLICANT: Tumas,Daniel
APPPLICANT: Watanabe, Colin K.
APPPLICANT: Williams,P.Mickey
APPPLICANT: Wood,WILLIAM I.
APPPLICANT: Zhang,Zemin
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: P2730PIC38
CURRENT APPLICATION NUMBER: US/09/997,653
CURRENT FILING DATE: 2001-11-15
PRIOR APPLICATION NUMBER: 60/049787
PRIOR FILING DATE: 1997-06-16
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/065186

[illegible]

```
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091478
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091544
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091519
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091626
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091633
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091978
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/091982
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/092182
; PRIOR FILING DATE: 1998-07-09

Query Match      3.88; Score 76.5; DB 9; Length 1184;
Best Local Similarity 21.28; Pred. No. 77;
Matches 72; Conservative 47; Mismatches 123; Indels 97; Gaps 19;

QY 17 YSIPVDNGVEGEPEIEGCPISITINENTRNAFEHGVY-----KGIYDGEGRNDE 67
      | : | | | | : : : : | : | : | : | : |
DB 660 FSYDFRDEVTSR-LNAGKVKVHLD-STQVKMPRHISTVKLWSLNPDTGLMEEGDKRFE 717
      | : | : | : | : | : | : | : | : | : |
QY 68 GGRVAGISLPFDSCNVAFTRSILNPRGIFVTTVVISFHPLEFYKYVD-----RAYR 118
      | : | : | : | : | : | : | : | : | : |
DB 718 NQRR-----NKRERT-----FLVGNLEIRERLFLNLDVESRRCYKVAAYR 760
      | : | : | : | : | : | : | : | : | : |
QY 119 VQCFYFMAADTVSKQIE---VSETTARFQIYPMYVC--RYELLDGGPTGQPVQPAIIG 173
      | : | : | : | : | : | : | : | : | : |
DB 761 SERE-----LPSEQIQGVIVISVINEPRGFLSNPRAMGRFSVITGPGACV----- 808
      | : | : | : | : | : | : | : | : | : |
QY 174 QPYVHKWTCDESEYVDFCAVHSCFVDDGNGDVEIL-----NAQSCALDKXLLNLTXY 227
      | : | : | : | : | : | : | : | : | : |
DB 809 -PAF-----CDDQSPDAYSAIVLASLA---GEEIQAVESPKKPNPAIGVPOPLYLNLT 859
      | : | : | : | : | : | : | : | : | : |
QY 228 -PTDLNAGQEAHVYKXADRSQLEFYQCOISITIKEPNSCEVRCPOCSEFQG--FGAVYTGGA 284
      | : | : | : | : | : | : | : | : | : |
DB 860 RRTD---HEDPRVKTKA-----FQISMAKRPNS-----AEEENGRIVAFENLRAC 902
      | : | : | : | : | : | : | : | : | : |
QY 285 AAKPAAAOURLKKSAPEENIIDVTDINTIEISDN 323
      | : | : | : | : | : | : | : | : | : |
DB 903 EEAAPSAAHFRFYQ-----IEGDRYDNTVPFEMDD 933
      | : | : | : | : | : | : | : | : | : |

RESULT 35
US-09-993-667-124
; Sequence 124, Application US/09993667
; Publication No. US20030022187A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi J.
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Geritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Napier, Mary A.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K.
```

```
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: P2730PIC4
; CURRENT APPLICATION NUMBER: US/09/993,667
; CURRENT FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: 60/049787
; PRIOR FILING DATE: 1997-06-16
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/065186
; PRIOR FILING DATE: 1997-11-12
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066770
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/075945
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/084600
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/087106
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087609
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087759
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087827
; PRIOR FILING DATE: 1998-06-03
; PRIOR APPLICATION NUMBER: 60/088021
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088025
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088026
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088028
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088029
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088030
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088033
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088326
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088167
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088202
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088212
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088217
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088655
; PRIOR FILING DATE: 1998-06-09
; PRIOR APPLICATION NUMBER: 60/088734
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088738
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088742
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088810
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088824
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088826
```


Query Match	3.8%	Score 76.5	DB 9	Length 1184
Best Local Similarity	21.2%	Pred. No. 77		
Matches 72	Conservative 47	Mismatches 123	Indels 97	Gaps 19
Query	17	YSIPDNCGVEGEPEIECGPTSTINFTRNAFEGHYV-----KGLYDEGGRNDE	67	
Db	660	FSVDFREVTSEP-LNMGKKYKVAHD-STQKKMEPHISTVAKLSLNDPTGLMEBEGCFKE	717	
Query	68	GGROYAGISLFPDSCNVAKTRSLNPRGIFVTTTVVISFHPLETKYD-----RAYR	118	
Db	718	NQRR-----NKRERT-----FLVGNLEIRERLFLVDPEBSRRCVYKRAYR	760	
Query	119	VQCYEMEDKTVSAQIE--VSELTTFQOIQVPMYC--RYELDGGPFGQVOPPAITG	173	
Db	761	SERF-----LPSEIOIGVIVISVNLBPRGFLSNPRAMGRFDSVTGPGACV-----	808	
Query	174	QPVYHKMTCDESEYDFECVAVHSCFVDGNGDRIEIL-----NADGALDKYLLNLEY	227	
Db	809	-PAF-----CDDQSPDAVSATVYLA SLA---GEELQAVESSPKRNPVAGVPOPYLKKLWY	859	
Query	228	-PTDLMAGQEAHYKVAADRSOLFQYCOISITIREPNSECVPOCSBPOG--FGAVTGA	284	
Db	860	RRTD--HEDPYVKTA-----PQISMAKRPNS-----AESNCPITAFELRLRC	902	
Query	285	AAKPAALAAQLRLKKRSAPENIIVRTDINTLEISDDN	323	
Db	903	EEAPPSAAHFRFYQ-----IEGDRYDYNTVTFPNEDD	933	
RESULT 40				
US-09-990-562-124				
Sequence 124, Application US/09990562				
Publication No. US20030027985A1				
GENERAL INFORMATION:				
APPLICANT: Ashkenazi, Avi J.				
APPLICANT: Baker, Kevin P.				
APPLICANT: Botstein, David				
APPLICANT: Desnoyers, Luc				
APPLICANT: Eaton, Dan L.				
APPLICANT: Ferrara, Napoleone				
APPLICANT: Fong, Sherman				
APPLICANT: Gerber, Hanspeter				
APPLICANT: Geriltsen, Mary E.				
APPLICANT: Goddard, Audrey				
APPLICANT: Godowski, Paul J.				
APPLICANT: Grimaldi, J. Christopher				
APPLICANT: Gurney, Austin L.				
APPLICANT: Kijavlin, Ivar J.				
APPLICANT: Napier, Mary A.				
APPLICANT: Pan, James				
APPLICANT: Paoni, Nicholas F.				
APPLICANT: Roy, Margaret Ann				
APPLICANT: Stewart, Timothy A.				
APPLICANT: Tumas, Daniel				
APPLICANT: Watanabe, Colin K.				
APPLICANT: Williams, P. Mickey				

```

; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P2730PIC18
; CURRENT APPLICATION NUMBER: US/09/990,562
; PRIOR APPLICATION NUMBER: 2001-11-14
; PRIOR FILING DATE: 1997-06-16
; PRIOR APPLICATION NUMBER: 60/049787
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/065186
; PRIOR FILING DATE: 1997-11-12
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066770
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/075945
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/084600
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/087106
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087609
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087759
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087827
; PRIOR FILING DATE: 1998-06-03
; PRIOR APPLICATION NUMBER: 60/088021
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088025
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088026
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088028
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088029
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088030
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088033
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088326
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088167
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088202
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088212
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088217
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088655
; PRIOR FILING DATE: 1998-06-09
; PRIOR APPLICATION NUMBER: 60/088734
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088738
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088742
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088810
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088824
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088826
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088858
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088861
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088876
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/0889105
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089440
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089512
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089514
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089532
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089538
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089598
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089599
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089600
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089653
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089907
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089908
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089947
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089948
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089952
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/090246
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090252
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090254
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090349
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090355
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090429
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090431
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090435
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090444
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090445
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090472
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090535
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090540
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090542
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090676
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090678
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090690
```


RESULT 2

T21239

hypothetical protein F22B5.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000

C:Accession: T21239

R:Slms, M.

submitted to the EMBL Data Library, July 1995

A:Reference number: Z19395

A:Accession: T21239

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-389 <MIL>

A:Cross-references: EMBL:Z50044; PIDN:CA90355.1; GSPDB:GN00020; CESP:F22B5.3

A:Experimental source: clone F22B5

C:Genetics:

A:Gene: CESP:F22B5.3

A:Map position: 2

A:Introns: 41/1; 116/2; 320/3

C:Superfamily: Caenorhabditis elegans hypothetical protein F22B5.3

Query Match 68.2%; Score 1382.5; DB 2; Length 389;
 Best Local Similarity 69.6%; Pred. No. 9.4e-114;
 Matches 268; Conservative 42; Mismatches 52; Indels 23; Gaps 8;

QY 12 LIALSYIFVNDGVEGEPEEGPSTITNFNTRNAFEGHVVYKGLYDGGCRNDGGRO 71
 DB 13 LVASVATFVNDNVEGEPEEGPSTITNFNTRNAFEGHVVYKGLYDGGCRNDGGRO 72
 QY 72 VAGISLPDSCNVAARTSLNPRGIFVTTTVISFHPLEFVKYDRAVRVOCFYMEADKTVS 131
 DB 73 VAGIELPFTCNVARTSLNPRGIFVTTTVISFHPLEFVKYDRAVRVOCFYMEADKTVS 132
 QY 132 AQIEVEITTAQTQIVPMPVCRYEILDGPGTPQVOPALIGOPVYHKMTCDSEYVDFTC 191
 DB 133 TQIEVSDLTAAQTQVMPVMPICKRYEILNGPTEGPVQFATIGQOVYHKMTCDSEYVDFTC 192
 QY 192 AVHSGFVNDGNDVVEILNADGALDKYLNNLEFPTDLMAQGEAHVYKYDRSOLFQV 251
 DB 193 AVHSGTVNDGNDVYQIILNDGALDKYLNNLEFPTDLMAQGEAHVYKYDRSOLFQV 252
 QY 252 COISTIKPEPNECVAPDCEPQEGFVAVTGGAAKPAAPAAQRLKKRSAR-PENIIVY 310
 DB 253 COISTIKPEPNECVAPDCEPQEGFVAVTGGAAKPAAPAAQRLKKRSAR-PENIIVY 307
 QY 311 RPDINTLEISDNOALPVLRHALL---QHNGCPVITLAAVONGICMSPFGS---MENG 364
 DB 308 RAEITTELELEGN-LPSSLTQAQALVASREIGD---SFRQELCISFHSISVTVFVG 361
 QY 365 LSTALIAAVIIT---ISFKFRPNOK 386
 DB 362 LTV--FVALFITVMTVSRMNPVPSDK 384

RESULT 3

A49772

cuticle protein cut-1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 01-Dec-2000

C:Accession: A49772; S27799

R:Sebastiano, M.; Lassandro, F.; Bazzicalupo, P.

Dev. Biol. 146, 519-530, 1991

A:Title: cut-1 a Caenorhabditis elegans gene coding for a dauer-specific noncollagenous

A:Reference number: A49772; MUID:91323673; PMID:1864469

A:Accession: A49772

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-423 <SEB>

A:Cross-references: GB:M55997

R:Sebastiano, M.; Lassandro, F.; Bazzicalupo, P.

submitted to the EMBL Data Library, July 1991

A:Description: CUT-1 a Caenorhabditis elegans gene coding for a dauer specific non co

A:Reference number: S27799

A:Accession: S27799

A:Molecule type: DNA

A:Residues: 'R', '77-78', 'YP', '121-423' <SE2>

A:Cross-references: EMBL:M55997; MID:g156271; PID:g156272

C:Genetics:

A:Gene: CUT-1

A:Introns: 360/3

C:Superfamily: Caenorhabditis elegans hypothetical protein F22B5.3

Query Match 68.1%; Score 1379; DB 2; Length 423;
 Best Local Similarity 65.8%; Pred. No. 2.2e-113;
 Matches 273; Conservative 31; Mismatches 67; Indels 44; Gaps 6;

QY 9 CTTLLALSTS-IPVDNGVEGEPEEGPSTITNFNTRNAFEGHVVYKGLYDGGCRNDE 67
 DB 8 CTAALVLSASAIIPVDNVEGEPEEGPSTITNFNTRNAFEGHVVYKGLYDGGCRNDE 67
 QY 68 GGRQVAGIELPFDSCNVAARTSLNPRGIFVTTTVISFHPLEFVKYDRAVRVOCFYMEAD 127
 DB 68 GGRQVAGIELPFDSCNVAARTSLNPRGIFVTTTVISFHPLEFVKYDRAVRVOCFYMEAD 127
 QY 128 KTVSAQIEVEITTAQTQIVPMPVCRYEILDGPGTPQVOPALIGOPVYHKMTCDSEYV 187
 DB 128 KTVSTQIEVSDLTAAQTQVMPVMPICKRYEILNGPTEGPVQFATIGQOVYHKMTCDSEYV 187
 QY 188 DFFCAVAVHSCFVNDGNDVVEILNADGALDKYLNNLEFPTDLMAQGEAHVYKYDRSOLFQ 247
 DB 188 DFFCAVAVHSCFVNDGNDVVEILNADGALDKYLNNLEFPTDLMAQGEAHVYKYDRSOLFQ 247
 QY 248 LFYQCOISTIKPEPNECVAPDCEPQEGFVAVTGGAAKPAAPAA----- 291
 DB 248 LFYQCOISTIKPEPNECVAPDCEPQEGFVAVTGGAAKPAAPAA----- 291
 QY 292 -----AQLRL-KKRS-APPENIIVDFPDINTLEISDNOALPV 328
 DB 308 APLVAAVFAAATAAFAVPRATLQRLRRKRSFGENGILDVREINTLDIMEG--ASPS 365
 QY 329 DLRRHALLDHNGCPVITLAAVONGICMSPFGSMKMLSTALIAAVITTSFRRP 383
 DB 366 APLAALVSE--ESVRRATSTGISSTPIGLPSFLDMITVATLATSATTFVAVP 418

RESULT 4

T22563

hypothetical protein F53F1.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T22563

R:Burton, J.

submitted to the EMBL Data Library, October 1996

A:Reference number: Z19583

A:Accession: T22563

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-364 <MIL>

A:Cross-references: EMBL:Z81088; PIDN:CA903124.1; GSPDB:GN00023; CESP:F53F1.1

A:Experimental source: clone F53F1

C:Genetics:

A:Gene: CESP:F53F1.1

A:Map position: 5

A:Introns: 57/1; 136/3; 186/3; 224/1

Query Match 35.8%; Score 725.5; DB 2; Length 364;
 Best Local Similarity 47.6%; Pred. No. 5.7e-56;
 Matches 129; Conservative 51; Mismatches 84; Indels 7; Gaps 1;

QY 19 IPVDNGVEGEPEEGPSTITNFNTRNAFEGHVVYKGLYDGGCRNDGGCRQVAGISLP 78
 DB 49 VPIQNSLIGDVVECDSTISVQITKEFVGVIFVKDPASEVCTSRGTGLSLAFLEIE 108
 QY 79 FDSCHVARTSLNPRGIFVTTTVISFHPLEFVKYDRAVRVOCFYMEADKTVSAQIEVSE 138

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 24, 2003, 08:50:03 ; Search time 19 Seconds
(without alignments)
1958.106 Million cell updates/sec

Title: US-10-054-562a-4

Perfect score: 2026

Sequence: 1 MMIRLAFCTTIALTSYIP.....ALIAVAITISFKRPNOKA 387

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_73:.*
1: pirl:.*
2: pirl:.*
3: pirl:.*
4: pirl:.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1420.5	70.1	424	2	T20032	hypothetical prote
2	1382.5	68.2	369	2	T21339	hypothetical prote
3	1379	68.1	423	2	A49772	cuticle protein cu
4	725.5	35.8	364	2	T22563	hypothetical prote
5	646	31.9	290	2	T24028	hypothetical prote
6	602.5	29.7	550	2	T23760	hypothetical prote
7	595	29.4	484	2	T20451	hypothetical prote
8	578.5	28.6	315	2	T21441	hypothetical prote
9	546.5	27.0	384	2	T24467	hypothetical prote
10	511	25.2	610	2	T22540	hypothetical prote
11	477.5	23.6	440	2	T24323	hypothetical prote
12	476	23.5	387	2	T26028	hypothetical prote
13	451	22.3	609	2	T25120	hypothetical prote
14	414	20.4	382	2	T27131	hypothetical prote
15	409.5	20.2	747	2	T34329	hypothetical prote
16	393	19.4	647	2	T23814	hypothetical prote
17	393	19.4	670	2	F88297	protein M28.1 (impr
18	376.5	18.6	477	2	T33556	hypothetical prote
19	357.5	17.6	437	2	T34211	hypothetical prote
20	339	16.7	225	2	T33114	hypothetical prote
21	293	14.5	366	2	T25042	hypothetical prote
22	290.5	14.3	344	2	T26228	hypothetical prote
23	289.5	14.3	1286	2	A88396	protein M01E10.2 (
24	151.5	7.5	514	2	T31736	hypothetical prote
25	141	7.0	3507	2	T34513	hypothetical prote
26	137	6.8	1262	2	T25168	hypothetical prote
27	132.5	6.5	665	2	T25228	hypothetical prote
28	120	5.9	1011	2	C87789	protein C34G6.6 (1
29	116	5.7	751	2	T21967	hypothetical prote

30	110.5	5.5	942	2 T19553	hypothetical prote
31	108	5.3	446	2 T19902	hypothetical prote
32	108	5.3	495	2 B89614	protein C43C3.3 (1
33	102	5.0	279	2 A83394	UMP-glucose-1-phos
34	101.5	5.0	828	2 T22367	hypothetical prote
35	100	4.9	935	2 S17855	peptidylglycine mo
36	94.5	4.7	875	1 URXLA2	submaxillary mucin
37	94.5	4.7	1569	2 T42233	hypothetical prote
38	93.5	4.6	649	2 T38883	hypothetical prote
39	92	4.5	2244	2 T11616	carbamoyl-phosphat
40	90	4.4	741	2 T22486	hypothetical prote
41	90	4.4	843	2 A40970	undulin 1 - human
42	90	4.4	1080	2 S48944	hypothetical prote
43	90	4.4	1197	2 T30581	neural cell adhesi
44	90	4.4	2515	2 A41519	posterior-group pr
45	88.5	4.4	1099	2 T14850	S-layer protein pr

ALIGNMENTS

```
RESULT 1
T20032
hypothetical protein C47G2.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
C:Accession: T20032
R:Palmer, S.
Submitted to the EMBL Data Library, April 1995
A:Reference number: Z19213
A:Accession: T20032
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1424 <N1>
A:Cross-references: EMBL:Z49125; PIDN:CAA8934.1; GSPDB:GN00020; CESP:C47G2.1
A:Experimental source: clone C47G2
C:Genetics:
A:Gene: CESP:C47G2.1
A:Map position: 2
A:Introns: 40/1, 115/2, 361/3
C:Superfamily: Caenorhabditis elegans hypothetical protein F2285.3

Query Match      70.1% Score 1420.5; DB 2; Length 424;
Best Local Similarity 67.1%; Pred. No. 4.8e-117;
Matches 279; Conservative 33; Mismatches 59; Indels 45; Gaps 7;

QY 9 CTTLIALSYS-IPVNGVEGEPEIEGPTSTTFNFRMAFEHVVYKGLYDGRNDE 67
      | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8 CLALVLSASAIIPVDNVEGEPEIEGPTSTTFNFRNPEGHVYKGLYDGRNDE 67

QY 68 GGRVAGISLPDSCNARTRSLNPRGIFVTTTVVIFSHFLFYTKVRAVRCFYMEAD 127
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 68 GGRVAGISLPDSCNARTRSLNPRGIFVTTTVVIFSHFLFYTKVRAVRCFYMEAD 127

QY 128 KTVSAOIESEITAFOTOLVPMVPCRYEILIDGSPGPOPAFATIGOPVYHKMTCDESETV 187
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 128 KTVSAOIESEITAFOTOLVPMVPCRYEILIDGSPGPOPAFATIGOPVYHKMTCDESETV 187

QY 188 DTFCAVHSCFVDGNDGDFEILNADSCALDKYLLNLMEYPTDLMAGOEAHVYKYADRSD 247
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 188 DTFCAVHSCFVDGNDGDFEILNADSCALDKYLLNLMEYPTDLMAGOEAHVYKYADRSD 247

QY 248 LFYQOCISITIKKEBNSCVPRQCSSEPOGFGAVKGTG-----AA 265
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 248 LFYQOCISITIKKEBNSCVPRQCSSEPOGFGAVKGTG-----AA 265

QY 286 AKPAAPVAAAAA--PAVPRATLQLRLRRKRSFGEENGLIDRVKVEINTLIDIEG--ASP 365
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 308 AKPAAPVAAAAA--PAVPRATLQLRLRRKRSFGEENGLIDRVKVEINTLIDIEG--ASP 365

QY 328 VDLRHRALDHOHNPVLAIVONGICMSPGFSFMGLSTALAAVITTSFKFRP 383
      | | : | : | | | | | | | | | | | | | | | | | | | | | | |
Db 366 SAPEAAALVSE--ESVRRRATSTGICLTPTIGFASFLGIGTIVATLSTATFYVARP 419
```


RESULT 6
T23760
hypothetical protein M142.2 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T23760
R:McMurray, A.
submitted to the EMBL Data Library, May 1996

[illegible]

```

QY 147 IVPMPVCRVETILOGSPFGPQVPAITIQPVPYHKKTCSEVVDTFCAVYHSCFVDDG-NGD 205
Db 132 ETQPLPVRYETILEMNA-GGSPIKYARIIDOVYHKKTCAEILENNYCMKHFCTVYDGGGP 190
QY 206 TVEILNADGALDKRYLLNNLEYPFDLMAAGQEAHYKRYKADDSOLFYOCCISITTKPEPSEC 265
Db 191 PTVIYDANCGSVDCVYIIQNIETYSIDLTAAGKLAPVFKFADKAGLTFNCQIOLITIKDVNYGC 250
QY 266 -VRPQCSEPO 274
Db 251 SNTOPQCPTSQ 261

RESULT 8
T21441
hypothetical protein F20D1.8 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T21441
R:Barton, J.
submitted to the EMBL Data Library, August 1996
A:Reference number: Z19382
A:Accession: T21441
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-315 <M1>
A:Cross-References: EMBL:Z78542; PIDN:CAB01742.1; GSPDB:GN00028; CESP:F20D1.8
A:Experimental source: clone F20D1
C:Genetics:
A:Gene: CESP:F20D1.8
A:Map position: X
A:introns: 32/3; 48/1; 80/2; 110/3; 158/2; 194/2; 245/1

Query Match 28.6%; Score 578.5; DB 2: Length 315;
Best Local Similarity 40.2%; Pred. No. 4,1e-43;
Matches 113; Conservative 47; Mismatches 76; Indels 45; Gaps 5

QY 4 RLIAFTTTLIAISYIPVDNGVEGEPEIECGPTITINFNTRNAFEGHYVYKGLYDOEGC 63
Db 30 RIPLPFCK-----YHAEQINGLOGEPLIRCGSESLINFKTGAFEGHYVYKGYSMKHC 84
QY 64 RNDEGRQVAGISLPEDSCNVARTRELNRGIFVTTTVVISHPLFVTVKDAIRYOCFY 123
Db 85 RFDATLESQVNLTVSYACDVIRORSSNPKGIMMTATIIISHPMEITRKIDSKYKOCFY 144
QY 124 MEADKTVSAQIEVS-----ETTAFOFOI----- 147
Db 145 AEMQKVTITQQLANVDIAKEBEKRIFWAVGDEEGYVSHITGDGKKLHKLNDPSTEERISTY 204
QY 148 VMPVCRVETILOGSPFGPQVPAITIQPVPYHKKTCDS---ETVDTFCAYVHSCFYDDGNG 204
Db 205 VPLPCKRYVLTRESKT-EEVAFATVGQIYVHEMSCGAPQNGTSPFCVYVHSCNNKDERG 263
QY 205 DRIVEILNADGALDKRYLLNNLEYPFDLMAAGQEAHYKRYKADR 245
Db 264 KEVQIFDENGCAVDKYLNNLEYSDDLDTGGLSOVCSTWVR 304

RESULT 9
T24467
hypothetical protein T04F8.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
C:Accession: T24467
R:Lenhard, N.
submitted to the EMBL Data Library, November 1995
A:Reference number: Z19895
A:Accession: T24467
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-384 <M1>
A:Cross-References: EMBL:Z66565; PIDN:CAA91480.1; GSPDB:GN00028; CESP:T04F8.4
A:Experimental source: clone T04F8

```

C:Genetics: A:Gene: CESP:T04F8.4
A:Map position: X
A:Introns: 16/3; 55/1; 99/3; 147/2; 195/1; 221/2; 276/3; 292/3
C:Superfamily: Caenorhabditis elegans hypothetical protein F22B5.3

Query Match 27.0%; Score 546.5; DB 2; Length 384;
Best Local Similarity 31.2%; Pred. No. 3.5e-40;
Matches 123; Conservative 75; Mismatches 155; Indels 41; Gaps 9

QY 5 LIAFCITLLALSTIPVNGVEGEPEIEGCPISITINENTRNAEFGHVVYKGLYDQGR 64
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 10 LVETFTVFLEFSAIIDDNDLIGOPETICNADIDMFRTKROFNKGKVSYNPRECR 69
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 65 -----DDEGROYAGSILPDCSNVARTSLNRGIFVTTVTIVSHPLFLTKRVARAYV 119
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 70 VDYSTDQGGRRPYVGKTLNHGACNMQRQMTAEGBMESTVLISHPLFTIMDAHYHI 129
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 120 QCFYMADKTVAQAIEVSEITTAFTQTQIYMPRCREIILDGPTGPVQAFATIGRPVHK 179
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 130 RCMYKEARTVTAIDVSNLPTEVSQSDLPMPICSYTIHRDDLDGIPLKYAVGQQVYHR 189
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 180 WTDDSETVDFCAVNVASCVDGNGCTVELLNADGCALDKYLNLNEYPTDL-MAGQEAH 238
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 190 WOODSE---DYGILLVHCYVEDQGCKOMIIDERGCHTRRLDGTPTYEALNNMAVRESF 246
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 239 YKKYARSOLOFYOCOSITIKPENSEC---VRPOCEPGFCFAVNTGGAAPAAQAOLR 295
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 247 VEFEADRIAVRQCETRLCLKD-DGGCDGITPPMCS-----FDANSENKOIYKRNR 298
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 296 LKKRSAGEENIIDVATDINTLTLEISPDONALPVDLNRHALOHNGOPVLAIVQNGICMS 355
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 299 TFRKPAD--WFSQVVMYMKENDGSTAQAAEIRE-----LDPTICTLA 341
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 356 P----FGSEMGLSTALIAAVITITSEREPNQ 385
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 342 PKLLVAVTFEFTEFIYLFVTTLLVVVYHRYCKNE 375
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

RESULT 10
T22540
hypothetical protein F53B6.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T22540
R:White, S.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z19578
A:Accession: T22540
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-610 <Will>
A:Cross-references: EMBL:Z81086; PIDN: CAB03116.1; GSPPDB: GN00019; CESP:F53B6.6
A:Experimental source: clone F53B6
C:Genetics:
A:Gene: CESP:F53B6.6
A:Map position: 1
A:Introns: 38/1; 103/1; 192/2; 239/1; 268/2; 300/3; 532/2; 546/1

Query Match 25.2%; Score 511; DB 2; Length 610;
Best Local Similarity 37.4%; Pred. No. 8.9e-37;
Matches 101; Conservative 58; Mismatches 101; Indels 10; Gaps 6;

QY 10 TTLALASTS--IPVNGVEGEPEIECGPTSTINENTRNAEFGHVVYKGLYDQGC--RN 65
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 61 TLILLSTSCFEIQNGGVGKPEVFCGIDTRLVKVNTHPHNRGIRLYDGSQKHCVQHS 120
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 66 DEGGQVAGISLPDSCNAVAKTRSLNRGIVTTVTVVISHPFLVTVKVDARYVQCYME 125
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 121 ADAHSSPQEFPIPIGACMMRKRTLHPRGISFSETMTISFPHFVLTGMDFASIRCFPLE 180
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 126 ADKTVSAQIEVSEITTAFTQTQIYMPRCREIILDGPTGPVQAFATIGRPVHKWTCDSE 185

QY 85 ARSLNDRGIEVTTTVVISHNPLFVYKVDRAHVQCFMEADKITYSAQIEVSEI--TTAF 143
Db 94 RRRNRVADPSTISYVTVIIQHHPLEVTSFKAAYKLNCIYQOESTLQORINVSIDISTAL 153
QY 144 QTOIVMPVRCRYEYELIDGGPTGQPOFAIIICOPVYHKWTCSETEVDFPCAVHSCFDDGN 203
Db 154 QSKNAPK--CRYDVLSSLSLNGPVYRFKRNVEDVYHNKTTCS---DREGFVHSCVVRDES 208
QY 204 GDVYEILNADSCALDKYLLNNLEYPYDLMAGQEA--HYUKYADRSQLFYQCSITITKEPN 262
Db 209 GKDRQFIDERCSCVDFESLFEVYSDDLKSAFTAVRAFRVADQVMHFSQITTCQKQEN 268
QY 263 S-ECVRRQCSBPQSGFAVK 280
Db 269 GCBGISPPICRPMDLGPIC 287

RESULT 14
 727131
 hypothetical protein Y53C12B.6 - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: J27131
 R:Kershaw, J.; Lennard, N.
 submitted to the EMBL Data Library, September 1997
 A:Reference number: Z20316
 A:Accession: J27131
 A:Status: Preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-382 <WIL>
 A:Cross-references: EMBL:Z299278; PIDD:CA816488.1; GSPDB:GN00020; CESP:Y53C12B.6
 C:Experimental source: clone Y53C12B
 C:Genetics:
 A:Gene: CESP:Y53C12B.6
 A:Map position: 2
 A:Introns: 23/1; 87/3; 167/1; 205/3; 279/3

[illegible]

RESULT 15
 T34329
 hypothetical protein K06A1.3 - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 C:Accession: T34329
 R:Fulton, L.
 submitted to the EMBL Data Library, July 1995
 A:Description: The sequence of *C. elegans* cosmid K06A1.
 A:Reference number: Z21507
 A:Accession: T34329

```

A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-747 <Full>
A:Cross-references: EMBL:023449; PIDN:AAC24298.1; GSPDB:GN00020; CESP:K06A1.3
A:Experimental source: strain Bristol N2; clone K06A1
C:Genetics:
A:Gene: CESP:K06A1.3
A:Map position: 2
A:Introns: 29/1; 62/3; 87/3; 133/3; 278/3; 430/3; 576/1; 659/3; 688/3

Query Match          20.2%  Score 409.5;  DB 2;  Length 747;
Best Local Similarity 31.4%;  Pred. No. 1e-27;
Matches 95;  Conservative 53;  Mismatches 112;  Indels 43;  Gaps 8

```

OY	6	I A F C T T L I A L S - - - - - I P V N G V G E B E I E D G P T S T I N P N R M A E G H V Y K G L X D O E	61
D b	4	I D F L L M L A L R N S Q P I Y E F P N H I S D E P E V S C H S G E M S L V K N E K S P P S H V F E K G H R R D	63
OY	62	G C R N D E G G R O V A G I S L P F D C S N V A R T S L N P R G I F V T T V I S E P H L P F V K D R A Y R V O C	121
D b	64	G C - - - S F S N T A N A T F E E S K C D V A R O R E A N P K M A T A Y A T V O U L H P L F T T K V D R A Y K L N C	119
OY	122	F Y M E A D K T V S A O I E - - - - - V S E T T A F T Q I Y P M P C R E I L D G P T Q	165
D b	120	F Y K E K V A V G A E V S R P E K A N N N N I T I K C S D P T P V O L D E S P O P C V S T Y I H K E S P N P	179
OY	166	P V O F A I I G P Y H R K W T C D S E T V D F F C A V H S C F E D D G N G D V I L A A D G A L D K Y L L N U	225
D b	180	I A K F A O L D V I Y A H W E C P S E - - - A Y Q E A V N C D V G E E Y S K R V I G E N G S E D I Y I M P L	236
OY	226	E Y P I D - I M A G E A H V Y K A D R S O L F Y O C O I S I - - T I K E R S E C V R Q C - - - - - S E	272
D b	237	I Y N E R K I K A F V N S A N F P D O N N V R I S C K I S V C P T L - - - S G T H O P K C D A S A S I D D A S E	293
OY	273	P O G	275
D b	294	P I G	296

RESULT 16
 T23814
 hypothetical protein M28.1 - *Caenorhabditis elegans*
 C.Species: *Caenorhabditis elegans*
 C.Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C.Accession: T23814
 R.Gardner, A.
 submitted to the EMBL Data Library, June 1995
 A.Reference number: Z19803
 A.Accession: T23814
 A.Status: preliminary; translated from GB/EMBL/DBJ
 A.Molecule type: DNA
 A.Residues: 1-647 <W1>
 A.Cross-references: EMBL:Z49911; PIDN:CAA50131.2; GSPDB:GN00020; CESP:M28.1
 A.Experimental source: clone M28
 C.Genetics:
 A.Gene: CESP:M28.1
 A.Map position: 2
 A.Introns: 17/3; 37/1; 226/2; 275/3; 323/2; 453/3; 540/3; 574/1

[illegible]

Db 379 HOCYVKGAGQOEVIDHOCCTLDQMLQTPYSEDSMSAODVAFKFPDRSTVDPRCT 438
QY 254 ISTTIKPNSECVRPOCSFQGGAVTGGAAKPPAAAOQLRLKKRSAPENIIDVRTD 313
Db 439 ITFCSDV-DAECL-----DMTPPKGANNMLRLRRKRSRSHGSLSLH 480
QY 314 INTLEI-----SDNOALPVDLRHRALLQHNQGPVILAAVQNGICMSPFQFEMGSL 366
Db 481 ANSLTYFDIDSSKTDITQMLPTPALLRKTEFEYERTECVASVSGILISASTF--FATIS 538
QY 367 IAL 369
Db 539 IAV 541

RESULT 17
F88297
protein M28.1 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C:Accession: F88297
R:Anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
A:Reference number: A75000; MID:99069613; PMID:9851916
A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C.elegans/
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A:Accession: F88297
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-670 <STO>
A:Cross-references: GB:chr_II; PIDN:CAA90131.1; PID:g3876695; GSPDB:GN00020; CESP:M28.1
C:Genetics:
A:Gene: M28.1
A:Map position: 2

Query Match 19.4%; Score 393; DB 2; Length 670;
Best Local Similarity 29.7%; Pred. No. 2.5e-26;
Matches 90; Conservative 56; Mismatches 127; Indels 30; Gaps 6;

QY 75 ISLPFSCNARTKSLNPRGIFVTVVISFHPLEVTKVDRAVVOCFYMEADKTVSAQI 134
Db 284 LEVPLNCSNTRKDRKLNPPSVVSLIAVSHFSITKLDKAYHICQAVAEAKTVSTDL 343
QY 135 EVSEITTAFOQIYPMVPCVCEYELIDGGPTGQPOFATIGQPYVYKTKCDSEVTFFCANY 194
Db 344 DVMMTDEQELNGTVPEPSCOTLISD--QNGSNVONSILVGLVRRQWCKGGLTKMLMLV 401
QY 195 HSCFVDDGNDGTVEILNADGCALDKYLLNNLEYPITLMAQO-EAHVYKVAQRSOLFQCO 253
Db 402 HOCYVKGAGQOEVIDHOCCTLDQMLQTPYSEDSMSAODVAFKFPDRSTVDPRCT 461
QY 254 ISTTIKPNSECVRPOCSFQGGAVTGGAAKPPAAAOQLRLKKRSAPENIIDVRTD 313
Db 462 ITFCSDV-DAECL-----DMTPPKGANNMLRLRRKRSRSHGSLSLH 503
QY 314 INTLEI-----SDNOALPVDLRHRALLQHNQGPVILAAVQNGICMSPFQFEMGSL 366
Db 504 ANSLTYFDIDSSKTDITQMLPTPALLRKTEFEYERTECVASVSGILISASTF--FATIS 561
QY 367 IAL 369
Db 562 IAV 564

RESULT 18
T33556
hypothetical protein W04C9.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T33556
R:Pulton, R.; Rohlfing, T.; Morris, M.
submitted to the EMBL Data Library, October 1998

A:Description: The sequence of C. elegans cosmid W04C9.
A:Reference number: Z21368
A:Accession: T33556
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-477 <FULL>
A:Cross-references: EMBL:AF098999; PIDN:AA68726.1; GSPDB:GN00019; CESP:W04C9.3
A:Experimental source: strain Bristol N2; clone W04C9
C:Genetics:
A:Gene: CESP:W04C9.3
A:Map position: 1
A:Introns: 10/3; 47/2; 140/2; 217/2; 300/1; 332/1; 458/2

Query Match 18.6%; Score 376.5; DB 2; Length 477;
Best Local Similarity 27.9%; Pred. No. 4.4e-25;
Matches 79; Conservative 65; Mismatches 118; Indels 21; Gaps 8;

QY 3 IRLIAFCTTLALSYSPVDNGVGEPEIECGPSTITNFTRNAFEGHYVVKGLYDOEG 62
Db 1 MKVLVLLMWNSTFSGDLNEITGEKRLQCNPSITFSFTRNPFIGNYIRGFYRTPG 60
QY 63 CR-----NDEGROVAGISLPDSCNVAARTSL--NPRGIFVTVVISFHPLEVTKV 113
Db 61 CRROFLAPNAGG---SFTVRLDCGMKRSRQISGSHSPRGVNOHITFVANFHNLTKE 116
QY 114 DRARVOCFYMEADKTVSAQIENVEI--TTAQTQIYMPVRCYRILLDGGPTGQPOFALI 172
Db 117 ERSNICKCFANSHSEGVKADIVSMPESEGVTVVTPQTYLRCEFTGEPKVSNTRV 176
QY 173 GOPYVHRKTCDESETVDTFCVAVVHSCFVDDGNG-DTVEILNADGCALDKYLLNNLEYPDL 231
Db 177 GMTIVHNMDCD--TSGNVGILIRCTILDSNGVSEFPLDDNGGSVSDPFQVYVLLSLT 234
QY 232 MAGQEAHYKKAQRSQLEFYOCQISITIKPNSEC---VRPCS 271
Db 235 SAYMAIEAISEFPDQPSISFSCQIKLCK-GSDRCRGMSPACT 276

RESULT 19
T34211
hypothetical protein F10E7.10 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34211
R:Paulley, A.
submitted to the EMBL Data Library, November 1995
A:Description: The sequence of C. elegans cosmid F10E7.
A:Reference number: Z21489
A:Accession: T34211
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-437 <PAU>
A:Cross-references: EMBL:U41264; PIDN:AAA82426.1; CESP:F10E7.10
C:Genetics:
A:Gene: CESP:F10E7.10
A:Introns: 41/3; 85/3; 266/3; 289/2; 328/2; 378/1

Query Match 17.6%; Score 357.5; DB 2; Length 437;
Best Local Similarity 31.6%; Pred. No. 1.8e-23;
Matches 87; Conservative 47; Mismatches 122; Indels 19; Gaps 7;

QY 1 MMIRLIAFCTTLALSYSPVDNGVGEPEIECGPSTITNFTRNAFEGHYVVKGLYDQ 60
Db 1 MLISIVWSFEFTIVVSLKYPNEIVDPPIVCEDEKRIIRIKTSMSPNSPIYVDEGND 60
QY 61 EGCNDEGROVAGISLPDSCNVAARTSLNPRGIFVTVVISFHPLEVTKVDRAVRVQ 120
Db 61 MECVS-----RNQNTIEVAHDKGCVVNEKTEQPNISIRLCIFVYDLHLPYTESRSYCAQ 116
QY 121 CFYMEADKTVSAQIEVS-EITTAFOQIIVP-----MPVCYREILLDGGPTGQPOFALI 173
Db 117 CVY--ADSHVWKDIESTLIDISEAPPLQSLSPQFADPVMKCVYSIRKQGNQGPVQVYASIG 174

[illegible]

Query Match	7.5%	Score 151.5	DB 2	Length 514
Best Local Similarity	19.0%	Pred. No. 3.1e-05		
Matches 67; Conservative	63;	Mismatches 108;	Indels 115;	Gaps 16;

RESULT 26
025168

RESULT 26
T25168
hypothetical protein T23F1.5 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T25168
R:Wilkinson, J.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z19990
A:Accession: T25168
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1262 <WILL>
A:Cross-references: EMBL:Z81129; PIDN:CAB03404.1; GSPDB:GN00023; CESP:T23F1.5
C:Experimental source: clone T23F1
C:Genetics:
A:Gene: CESP:T23F1.5
A:Map position: 5
A:Introns: 49/3; 201/1; 505/3; 609/3; 909/1; 944/2; 1025/2; 1079/1; 1168/3; 1225/1

Query Match 6.8%; Score 137; DB 2; Length 1262;
Best Local Similarity 20.8%; Pred. No. 0.002;
Matches 80; Conservative 60; Mismatches 188; Indels 56; Gaps 15;

OY 21 VUNGVGEPEIEICGPTSTINENTRNAEEGHVYAGLYDQEGCRNDGGGRVAGISLPED 80
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 903 VNRRIGKPRILCLEDEGISFEVKITLPLTGEVFANDRKREKLCHKSFTBEAKPKMELPFS 962
OY 81 SCNVART-RSLNPGIFVTTVVISPHPLFVTKVD-----RAVRCCFMEND---- 127
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 963 TCQVKNKGQDVSR-----GYHMOVLIIQNGNSTNLQSMACCVHQKKVYNKQ 1013
OY 128 ---KTVASQAEVSEITTAFOTQIVPMPCRYEI-LDG--PTGORPOFAIIGOPYHNKWT 181
Db 1014 VLPRKEALELELRILVPYSKLEQGKASMPVSVMQIIVDEGHHKLGAEVSAADIMPLALKWS 1073
OY 182 CSSEVDTFCANVHSCFVDNGDIVE-ILNADGALDKYLINLETPP-DLMAGOEAHV 239
Db 1074 MWPEE-DAAGMHIRCKKVDAVGKIDHTLLIDEQGSADQLIIDHPHYDTYHDITASHHMA 1132
OY 240 YKYARSOIFYCOQISTITKEPNSECVRPQCSEPQGFVAVKTGAAMAPAAQAOLRLKK 299
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 1133 FVPPMSSLIQIKCDILICSNIKSVYNTTSCED-----IPSPFCADV----- 1175
OY 300 RSAFEENIT-DVRTDINTLETISDDNQALPV--DLRHALLQHNGOPVILIAVQNCICMS 356
Db 1176 VTSPPNISDSASPFIKRHRASVPSEVSPATQSVTSTICLSRSCTPDPSEVVR--TCVNT 1233
OY 357 GFESMFMGSLIALIAAVITTTISFK 380
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 1234 QLATTSTGLSIAFL--LFATSFQ 1254

RESULT 27
T25228
hypothetical protein T24C2.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
C:Accession: T25228
R:Baynes, C.
submitted to the EMBL Data Library, November 1995
A:Reference number: Z20000
A:Accession: T25228
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-665 <WILL>
A:Cross-references: EMBL:Z66120; PIDN:CAA92199.1; GSPDB:GN00028; CESP:T24C2.1
C:Experimental source: clone T24C2
C:Genetics:
A:Gene: CESP:T24C2.1
A:Map position: X
A:Introns: 31/1; 57/3; 95/3; 141/2; 175/2; 196/3; 226/1; 259/1; 328/3; 388/2; 431/2; 522/3
C:Superfamily: Caenorhabditis elegans hypothetical protein T24C2.1

Query Match 6.5%; Score 132.5; DB 2; Length 665;
Best Local Similarity 27.6%; Pred. No. 0.0021;
Matches 35; Conservative 18; Mismatches 69; Indels 5; Gaps 2;

```

QY 5 LIACTTLLATXSYPVNDG--VEGEPEIECGPTSTITINENRMAEGHVYKGLYDQGC 63
      | : : : : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 8 VVGITFSFVVLATSTDFEDNAHMGVPQVYCSAKLLITVFSFNINIPQGISVFDKLFIPAC 67
      | : : : : : | : : | : : | : : | : : | : : | : : | : : | : : |
QY 64 RNDEGGROYAGISLPFSDSCNVARTRSLNPRGIFVTTTVVISFHPLEVTYKVDRAVRVOCFY 123
      | : : : : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 68 NHDYSTNIQKNATFQMDILKCANPMFLKNGSHLLRAYEIGHPLVMYNSDRFLVECL- 126
      | : : : : : | : : | : : | : : | : : | : : | : : | : : | : : |
QY 124 MEADKTV 130
      | : : : : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 127 ---DNTI 130

RESULT 28
C87789
protein C34G6.6 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C:Accession: C87789
R:Anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating bio
A:Reference number: A5000; MUID:99069613; PMID:9851916
A>Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/projects/C_
A:Accession: published ereta appeared in Science 283, 35, 1999; Science 283, 2103, 1999;
A:Accession: C87789
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1011 <STO>
A:Cross-references: GB:chr_I; PID:AB52479.1; PID:g1943841; GSPDB:GN00019; CESP:C34G
A:Gene: C34G6.6
A:Map position: 1

Query Match 5.9%, Score 120; DB 2; Length 1011;
Best Local Similarity 18.3%, Pred. No. 0.047;
Matches 84; Conservative 77; Mismatches 173; Indels 124; Gaps 25;

QY 18 SIPVNGVEGEPE--IECGPTSTITINENTNAFEGHYVYGLYDQGCGR--NDEGGR 70
      | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 577 TVPIPPVYVKAKEALIQITICNTEGIKVQINCEPESGVLEVNKKFDY--CRVEYANSNAAT 634
      | : : : : : | : : | : : | : : | : : | : : | : : | : : | : : |
QY 71 QVAGISLPF-----DSCNVARTRS----- 89
      | : | : : : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 635 IYLGLPKDFGMRPISLDIINDNENGTKNKTKKGEPLKDEIEFRQKROAAEPRDGLVD 694
      | : : : : : | : : | : : | : : | : : | : : | : : | : : | : : |
QY 90 -LNPICITVTTTVVISFH---PLEVTKYDRAVRVOCFMEADKTVSAQIEYS---EITTA 142
      | : : : : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 695 LLN--GYKSTVVIOTNNMIGIPGLVTSMDQLYEVSCDY---SSMGLGRVAGVNMVTGP 749
      | : : : : : | : : | : : | : : | : : | : : | : : | : : | : : |
QY 143 FQTOIVPM-----PVCRYEILIDGPGTGPQVQVGAIIIGQAPVYHWW---TQDSEVDFECA 192
      | : : : : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 750 EANLIQPRGKIELGNPLY--MQLNGDGTBPLOYAKIGDILIELMWTAMADDE-LDFE-- 805
      | : : : : : | : : | : : | : : | : : | : : | : : | : : | : : |
QY 193 VVHSCFVVD-----GNGDTVEILNADGC----ALDKYLLNNLEYPTDLMAQGEAHYK 241
      | : | : : : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 806 -VKNHARPGVAGKAGAGEKRLRIDG--GCTPAVAKKILIPGALIEISSAVKTTKMAFR 863
      | : : : : : | : : | : : | : : | : : | : : | : : | : : | : : |
QY 242 YADNSQLFYQCQIISITTKEPNSCEVRQCSPEQSGFAGVKTGGAAPAAQAOLR----- 295
      | : : : : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 864 FDDSSASIVTCEVEIC----KGDEPVECALT---GGVKKSFGHKRREVSNNIEEFETNR 916
      | : : : : : | : : | : : | : : | : : | : : | : : | : : | : : |
QY 296 -LLKRSAPENNI-----DYRDIINTLEISDNQALPVDLRRA---LLQNGQPVIL 345
      | : : : : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 917 YLIPRSATTSIYIIDPLQOVCTKVN-----PVAMSRASTLDLLREDAHEVOM 966
      | : : : : : | : : | : : | : : | : : | : : | : : | : : | : : |
QY 346 AAVQNGICM--SPFGFSMEGLSIALIAAVITTSFKER 382
      | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 967 IE-EGSICLNSVTVFAIFGLTAVILIGQTVIAHYAIVR 1003
      | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |

```


C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Nov-2001
 C/Accession: B89614
 R/Anonymous: The C. elegans Sequencing Consortium.
 C:Science 282, 2012-2018, 1998
 A>Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
 A:Reference number: A75000; MUID:99069613; PMID:9851916
 A>Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_ele
 A>Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
 A:Accession: B89614
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-495 <STO>
 A:Cross-references: GB:chr.X: PIDN:CAA87330.1; PID:q3874884; GSPDB:GN00028; CESP:C43C3.3
 A>Note: CDNA EST EMBL:T01034 comes from this gene
 C/Genetics:
 A:Gene: C43C3.3
 A:Map position: X

Query Match 5.3%; Score 108; DB 2; Length 495;
 Best Local Similarity 22.6%; Pred. No. 0.2; Mismatches 128; Indels 88; Gaps 19;
 Matches 77; Conservative 47;

QY 31 IECGPTSTINFN-----TRNAFGHYVKGLYDOEGC---RNDEGGRQVAGISL 77
 DB 83 VDCIADSFVYVINKSDPEYMRMISNPKSQPVYVYGHKTRHPCGTSKMEKGLTNFNLT 142
 QY 78 PPDG-CNVARTRSLNRCIGFTTVVISFH-PLEVTKVDRAVRVOCFYNEADKTVSAQIE 135
 DB 143 PYGSECDVLTDL--PKHRYAETTVYLEDNADLSFGKTRRLNHVFLCYLRANKT-----IR 196
 QY 136 VSEITTAFC---TQIVPVPVCRYEL---DGGPTGQP-----VQFIIIGQP--VYHK 179
 DB 197 FSDVSNGHVIVASTGCKPKP--KVELFRSTDSGKTLQAARENFEFEFIALSPDSAVHG 254
 QY 180 WT-----CDSEYVD-----TFCAVYHSCFVDDGNGDTVEILNADGALKILNNNE 226
 DB 255 ISPKRECTESDRDISAPPAKKITF--VGCGCPYNGMND-----IIDPLA 296
 QY 227 YPTDLMAQEAHVYKADRSOLFYOQOISITIKEPNSECVRPQSGPQGVAKTGGA 286
 DB 297 NVNDQIVFSKFRFRFGNOSTVFHCQVQVCLK--DECSK--TCYKK-----VSD 343
 QY 287 KPAAAOGLRLKKRSAPENITIDVKTINTLEISDDNAL 326
 DB 344 SNLTAERLKFRHRS-----ITDLERKTRTSAPATDNGSL 378

RESULT 33

A83394
 UTP-glucose-1-phosphate uridylyltransferase PA2023 [imported] - Pseudomonas aeruginosa
 C:Species: Pseudomonas aeruginosa
 C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 C/Accession: A83394
 R/Stover, C.K.; Pham, X.O.; Eryin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Br
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; Lim,
 J.; Lory, S.; Olson, M.V.
 A>Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
 A:Reference number: A82950; MUID:20437337; PMID:10984043
 A:Accession: A83394
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-279 <STO>
 A:Cross-references: GB:AE004628; GB:AE004091; NID:99948018; PIDN:AA05411.1; GSPDB:GN001
 A:Experimental source: strain PA01
 C/Genetics:
 A:Gene: galU; PA2023
 C:Superfamily: Escherichia coli utp-glucose-1-phosphate uridylyltransferase

Query Match 5.0%; Score 102; DB 2; Length 279;
 Best Local Similarity 21.9%; Pred. No. 0.3;
 Matches 49; Conservative 32; Mismatches 89; Indels 54; Gaps 11;

QY 28 EPEIECGPTSI-TINENTRNAFGHYVKGLYDOE-GCRNDEGGRQVAGISLPEFDCNVA 85
 DB 41 EEALFAGLSEIGIVTGRKRSLEDHFDIS--VELEHQIRNTDMEKYLVRRLIDECTFA 98
 QY 86 RTFSLNPRGI---FVTTTVVISFHLPLFYKVDRAVRVOCFYNEADKTVSAQIEVSITTA 142
 DB 99 YTRQVMEKGLGHAITLGRPLIDGEPVAVLADL---CLNLEGDSVLKQW---KLYNQ 151
 QY 143 FQTOIV-----PMVCRREIILDG-----GPTGQPVQFAIGQPVYHK 179
 DB 152 FRCSIVAIDVPEEENKKGVLAGEMIRDDIRVNTWMEKRPKEARPSMALIIGRIYL-- 209
 QY 180 WTCDSFTVDFCAVYHSCFVDDGNGDTVEILNA-----DGCAL 217
 DB 210 -----TPDIFDLIEQT---EPKGGEIQTIDALMKQADGCVL 244

RESULT 34

T2367
 hypothetical protein F47G9.3 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
 C/Accession: T2367; T24409
 R.White, S.
 submitted to the EMBL Data Library, June 1996
 A:Reference number: T2367
 A:Accession: T2367
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-828 <W12>
 A:Cross-references: EMBL:Z74035; PIDN:CAA98485.1; GSPDB:GN00023; CESP:F47G9.3
 A:Experimental source: EMBL:Z74041; PIDN:CAA98523.1; GSPDB:GN00023; CESP:F47G9.3
 R.Mortimore, B.
 submitted to the EMBL Data Library, June 1996
 A:Reference number: Z19887
 A:Accession: T24409
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-828 <W12>
 A:Cross-references: EMBL:Z74041; PIDN:CAA98523.1; GSPDB:GN00023; CESP:F47G9.3
 A:Experimental source: clone T03F7
 C/Genetics:
 A:Gene: CESP:F47G9.3
 A:Map position: 5
 A:introns: 62/1; 181/1; 375/2; 420/1; 452/1; 585/2; 662/2; 740/3

Query Match 5.0%; Score 101.5; DB 2; Length 828;

Best Local Similarity 20.6%; Pred. No. 1.5; Mismatches 181; Indels 67; Gaps 18;

Matches 80; Conservative 61;
 QY 18 SIPVNGVGEPEIECGPTSTI---NENTRNAFGHYVKGLYDOEGCRNDEG--GRQV 72
 DB 464 SSPLD--VKISQCTGFGFISVYSPGNPDNHLVAGGEIYVRN--GHSNCSKITIGDGERI 519
 QY 73 AGISLPEDSCNVARTRSLNRCIGFTTVVISFHLPLFYKVDRAVRVOCFYNEADKTVSA 132
 DB 520 LKIRNDPTTC-ITKNGDIEYFVVVTVQNVESGNAFTVITIDQLFKVCDSNQNNAVAV 578
 QY 133 QIEVSEITTAFC---TQIVPVPV---CYEILLDGGPTGQPVQFALIGQPVYHKKTC 182
 DB 579 AKTMRLRTTOFNKLDIYKVVNPKPMSMELRGREIT-----KARTVKLGSDLVETV 631
 QY 183 DSEYVDFCAVYHSCFVDDGNGDTVEILNADGCALD---KYLNN--LEYPTDLMAQEA 237
 DB 632 DNST-SAHNVYVKKCTAVDQNGDEKILIKNGCAQNAKEVLRDEIKETATGFLPPRA 690
 QY 238 HVYKADRSOLFYOQOISITIKEPNSECVRPQSGPQGVAKT--GGAAPKAPAAAOGLR 295
 DB 691 FRFKGGAVKI--EEVAYCKE-----CKKVRFFVMIKNYFGNIIFN-----N 731
 QY 296 LLKRSAPENITIDVKTINTLEISDDNAL---PVDLRHALLQHNQOPVILAAYONGI 352
 DB 732 LLDNATINIANCPSRNRFTTFDESEDTTLTGNEDETEKVAHELLV---QTIVDSYOLDGY 788

QY 242 YADRSQILFYQCQISTITKEPNSCVRPQCSPPGCGAV 279
 Db 1442 -----EYGASPADPNNPCISYCHN-TGFEVAV 1467

RESULT 38
 T38883
 hypothetical protein SPAC4H3.03c - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jun-2000
 C:Accession: T38883
 R:Murphy, L.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
 submitted to the EMBL Data Library, February 1996
 A:Reference number: Z21807
 A:Accession: T38883
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-649 <MUR>
 A:Cross-references: EMBL:269380; PIDN:CAA9342.1; GSPDB:GN00066; SPDB:SPAC4H3.03c
 A:Experimental source: strain 972h-; cosmid c4H3
 C:Genetics:
 A:Gene: SPDB:SPAC4H3.03c
 A:Map position: 1
 C:Superfamily: Schizosaccharomyces pombe hypothetical protein SPAC4H3.03c

Query Match 4.6%; Score 93.5; DB 2; Length 649;
 Best Local Similarity 22.1%; Pred. No. 5.4;
 Matches 86; Conservative 52; Mismatches 145; Indels 107; Gaps 21;

QY 25 VEEPEIEGCPITITNTNTRNF-----EGHVYVKGILYDQEGCRNDEGGHQAAGISL 77
 Db 36 LDGSEVMKMP-----NFSPSIFARILDARAGHFSITPI-EQTQCK-----QK 78

QY 78 PFDSCVARTRSINPRGIEFTTVVISF-----PLEVTKYDRA-----YRVQCF 122
 Db 79 YESTNIMLHKFKYSEKGVL---RLDFFHPRMEDYERPLMLIRKYSCTIRGTSRIKLECG 135

QY 123 YMEADTVSAQIEVSEIT-TAFQGTQVPMPCRYEILDGGPTQ-----PYQFAI 171
 Db 136 PALDYARQSHETRVSKITEVYQAEFVPASGDPKYLIDCVPSGQDKIDLELILPRAHNI 195

QY 172 IGQPVYHKMTCSEYVDTCFAVHSCFVDDGNDVYEILNADGALDKYILNNLEYPTDI 231
 Db 196 EGGGVLSYLELEGGQITF-----TFQEGGLDPNDVYTPN-----LVKLELDSKTR 242

QY 232 MAGQEAHVYKRYADRSQILFYQ---CQISTITKEPNSCV-RPQCSPPGCGAVKT----- 281
 Db 243 YMRWAIQOCVYIGRYREFYQARNALTKLILYPTGAVIASPTSPEDLGAVANNMDYRFT 302

QY 282 --GGAARPAQAQ-----RLAKRSAREP-NIIDVTRDINTLEISDDNQ 324
 Db 303 WTRDSAFITTYLAQLGFRAEAVEYMSFIYVLKKKKDGINIV-----YSHIGDSQ 354

QY 325 AL-PVDLRH-RALDHNGQPVIL--AAVON 350
 Db 355 NLEVEVLTLHRC--YINSHPRIGNAVAH 382

RESULT 39
 T11616
 carbamoyl-phosphate synthase (glutamine-hydrolyzing) (EC 6.3.5.5) - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jan-2000
 C:Accession: T11616
 R:Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.; Wood, V.
 submitted to the EMBL Data Library, October 1995
 A:Reference number: Z17300
 A:Accession: T11616
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-2244 <BAR>
 A:Cross-references: EMBL:254328; NID:g1009451; PID:g1009456

A:Experimental source: strain 972h(-)
 C:Genetics:
 A:Map position: IL
 A:Note: SPAC22G7.06c
 C:Superfamily: rudimentary enzyme; aspartate/ornithine carbamoyltransferase homology; arbamoyl-phosphate synthase (glutamine-hydrolyzing) large chain homology; carbamoyl-P F;61-1519/Domain: carbamoyl-phosphate synthase (ammonia) homology <CPA>
 C:Keywords: ligase
 F;61-440/Domain: carbamoyl-phosphate synthase (glutamine-hydrolyzing) small chain hom F;265-440/Domain: trpG homology <TRG>
 F;476-1516/Domain: carbamoyl-phosphate synthase (glutamine-hydrolyzing) large chain h F;476-926/Domain: biotin carboxylase homology <BCB>
 F;1532-1857/Domain: Bacillus dihydroorotase homology <DHO>
 F;1937-2237/Domain: aspartate/ornithine carbamoyltransferase homology <AOC>

Query Match 4.5%; Score 92; DB 2; Length 2244;
 Best Local Similarity 20.3%; Pred. No. 41;
 Matches 74; Conservative 58; Mismatches 153; Indels 80; Gaps 17;

QY 42 FTRNAFEHGHVYVKGILYDQEGCRNDEG-----GRQVAGISLPFDSCNVARTRSLNPRGI 95
 Db 999 YTYNNAVEHDHF-----NDKGVWVLGSGVYRIGSSVFEFDMCAVRAVETLDRGV 1048

QY 96 FTTTVVISFHPLEVT---KYDRAYVQCFYMEADKTVSAQIEVSEITTAFOQT---V 148
 Db 1049 ---KITMVNVPETVSTDYDEADRLY-FENIGLETVLIDYEGSSGIIIMAGGQTANNI 1104

QY 149 PMPVYCV-----EIIDGGT---GQPVFAIIQPVYHKMTCSEYVDTCFAVH 195
 Db 1105 ALPLHRENVKILGTSPKIDGAENRRKFSRMLDIDGVDPKKMLTSFDE-ADKFC---- 1159

QY 196 SCFVDDGNDTV---EILNADGALDKYILNNLEYPTDI-----MAGQEAH---YKYA 243
 Db 1160 -----DYGYPVLVPRSYVLGSAANNYVSGDLSYQQAIVAIKNDHPVYSKITI 1210

QY 244 DNSQILFYQCQISTITKEPNSCVRPQCSPPGCGAVKTGAA---AKPAAQAQRLKK 299
 Db 1211 ENAK-----EIEDAVAREGKMVHVISHEVENAGVHSGDATLVLPQDLATFTTERIVD 1265

QY 300 RSEAPENIIDVTRDINTLEISDDNQALPVDLHRALDHNGQPVILAAVONICSPGCF 359
 Db 1266 AAKIGSEALNITGPYITQFIARDNEIKVIECNVA---SRSPFVSKVI--GVDMISMAT 1320

QY 360 SMFMG 364
 Db 1321 DVIWG 1325

RESULT 40
 T22486
 hypothetical protein F52B11.3 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T22486
 R:Matthews, L.
 submitted to the EMBL Data Library, November 1996
 A:Reference number: Z19569
 A:Accession: T22486
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-741 <WII>
 A:Cross-references: EMBL:282268; PIDN:CAE05199.1; GSPDB:GN00022; CESP:F52B11.3
 A:Experimental source: clone F52B11
 C:Genetics:
 A:Gene: CESP:F52B11.3
 A:Map position: 4
 A:Introns: 48/1; 298/1; 341/3; 690/2

Query Match 4.4%; Score 90; DB 2; Length 741;
 Best Local Similarity 22.1%; Pred. No. 13;
 Matches 89; Conservative 43; Mismatches 181; Indels 90; Gaps 20;

QY 22 DNGVEGEPEIE-----CGPTSTITNTNTRNFAEGHVYVKGILYDQEGCRN- 65

```
Db 366 DANIKANPSMDYEFNCESGFGMALCTNEGIRFIVNTKEPYTGAIYAAERFST--CSQV 423
QY 66 DEGRQVAGTSLPPDSQCNVATRSLNPRGIVTTTVISFHLF---VTKVDRAIYQC 121
Db 424 VENAKQ--ISTTPPTVSSDCGTIIRDGKMEALVWVSLDGLPHOVTTMWRFRYRVC 480
QY 122 -----FYMEADKTVSAQIEVSEITTFQTOIVMPV-CRYEILDGPTGQPVQFAITGQ 174
Db 481 DVSMKMKVKEGSSVYVTTIYEASSQNTVLDVATPPVSAELQILN--QLEPPLIKASIGD 538
QY 175 PVYHKWTCDESETVDFCAVHSCFVD--DGNQDVEI-LNADGCALDKYLLNMLEYPTDL 231
Db 539 PLLIVIT--SEQAGPHNMWVTECTATRVGFGDVTVPFTLIENGCP-----RYPA-L 586
QY 232 MAGQEAHVYKADRSQI-----FYQCQISITIKERNSECVRPQCSFPOGFGAVKTGGA 284
Db 587 VGPVEQDFDKNRILKSDLRARFLDGSYDQIVCSIM-----FCAGPNCGPVSNCLDS 637
QY 285 AAKPAAQAQLRLKKRSAEPENIIDVPTDINTLEISDDNQALPYDLR-----HRALL 336
Db 638 GTNELFMHGR--KKRSAD-----LANGETEEKLSAIIIRVFAKGEDEDEMEM 682
QY 337 QHNGQPVILAAVQNGICMS--PEGFSMFMGLSIALIA-AVIITI 377
Db 683 ANNTMTMSMSDSTELICIAEPFVSVVSLVLCFALSAITAI 725
```

Search completed: March 24, 2003, 08:51:05
Job time : 30 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 24, 2003, 08:49:12 ; Search time 13 Seconds

(without alignments)
1234.718 Million cell updates/sec

Title: US-10-054-562a-4

Perfect score: 2026

Sequence: 1 MMRLIAFCTLIATLSYSIP.....ALIAAVITTSKFRPNQKA 387

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1420.5	70.1	424	1	CUTL1_CAEEL
2	108	5.3	446	1	YOH3_CAEEL
3	102	5.0	279	1	GAU1_PSEAE
4	96	4.7	279	1	GAU2_PSEAE
5	94.5	4.7	875	1	AMD2_XENLA
6	93.5	4.6	649	1	YAY3_SCHPO
7	92	4.5	2244	1	PYR1_SCHPO
8	90	4.4	1080	1	NRK1_YEAST
9	90	4.4	2515	1	TUD1_DROME
10	88	4.3	429	1	THRC_BUCAT
11	86.5	4.3	407	1	CGEL_CHICK
12	86.5	4.3	1260	1	YAOE_SCHPO
13	86	4.2	404	1	ISCS_METTE
14	85.5	4.2	1117	1	DSG2_HUMAN
15	85	4.2	1025	1	HIRA_FUGRU
16	85	4.2	4485	1	DYHG_CHIRE
17	84.5	4.2	410	1	CAOC_CANFA
18	84	4.1	2193	1	POL6_EC12T
19	83.5	4.1	563	1	MUC6_BOVIN
20	83	4.1	666	1	TKT1_BACHD
21	83	4.1	820	1	PDZ1_ARATH
22	83	4.1	1627	1	PAPA_HUMAN
23	81.5	4.0	430	1	SYN_BACHD
24	81.5	4.0	2286	1	DPOE_HUMAN
25	81	4.0	848	1	TORAECOLI
26	81	4.0	1080	1	CARB_XYLA
27	81	4.0	1663	1	CO3_MOUSE
28	80.5	4.0	311	1	RLAO_CAEEL
29	80.5	4.0	430	1	SYN_STAM
30	80.5	4.0	864	1	LOX2_PEA
31	80	3.9	848	1	TORAECOLI
32	80	3.9	974	1	AMD_HUMAN
33	79.5	3.9	340	1	PURA_METTH

34	79.5	3.9	1753	1	YEW2_YEAST	P32634 saccharomyc
35	79.5	3.9	2214	1	PYR1_YEAST	P07259 saccharomyc
36	79	3.9	553	1	YF61_METJA	O58956 methanococc
37	79	3.9	750	1	CTPB_MYCLE	P46840 mycobacteri
38	79	3.9	827	1	CADH_MOUSE	O91100 mus musculi
39	79	3.9	3587	1	TYCB_BACBR	O30408 b tyrocidin
40	78.5	3.9	299	1	RNHL_HUMAN	O75792 homo sapien
41	78.5	3.9	436	1	RP54_BACSU	P24219 bacillus su
42	78.5	3.9	556	1	GLI_CHICK	P55878 gallus gall
43	78.5	3.9	570	1	ZPR_FELCA	P48834 felis silve
44	78	3.8	507	1	RREP_MEASI	P26033 measles vir
45	78	3.8	785	1	SM3E_CHICK	O42237 gallus gall

ALIGNMENTS

RESULT 1	ID	CUTL1_CAEEL	STANDARD:	PRT:	424 AA.
AC	003755	Q18693:			
DT	01-FEB-1994	(Rel. 28, Created)			
DT	15-JUL-1998	(Rel. 36, Last sequence update)			
DT	15-DEC-1998	(Rel. 37, Last annotation update)			
DE	Cuticlin 1 precursor.				
GN	CUT-1 OR CATG2.1.				
OS	Caenorhabditis elegans.				
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;				
OC	Rhabditidae; Pelodierinae; Caenorhabditis.				
OX	NCBI_Taxid=6239;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-Bristol N2;				
RX	MEDLINE=91323673; PubMed=1864469;				
RA	Sebastiano M., Lassarano F., Bazzicalupo P.;				
RT	"cut-1 a Caenorhabditis elegans gene coding for a dauer-specific				
RT	noncollagenous component of the cuticle.";				
RL	Dev. Biol. 146:519-530(1991).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-Bristol N2;				
RA	Palmer S.;				
RL	Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.				
CC	-1- FUNCTION: COMPONENT OF THE CUTICLES. IT FORMS A RIBBON				
CC	APPROXIMATELY 2 MICRONS WIDE RUNNING ALONG THE LATERAL LINES				
CC	UNDERNEATH THE ALAE. CONTRIBUTES TO THE FORMATION OF EXTRACELLULAR				
CC	ENVELOPES PROTECTING THE ORGANISM FROM THE ENVIRONMENT.				
CC	-1- SUBCELLULAR LOCATION: Localized in a narrow band not exposed				
CC	TO THE EXTERIOR. RUNNING UNDERNEATH THE ALAE.				
CC	-1- DEVELOPMENTAL STAGE: EXPRESSED DURING DAUER LARVA FORMATION.				
CC	-1- DOMAIN: THE SMALL REPEATS A-A-P-[AVI] ARE ALSO PRESENT IN MANY				
CC	PROTEINS CONSTITUTING THE PROTECTIVE ENVELOPE OF OTHER SPECIES.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL: M55997; AAA27995.1; ALT_INIT.				
DR	EMBL: Z49125; CAA88934.1; -.				
DR	PIR: A49772; A49772.				
DR	WormPep: C47G2.1; CE02165.				
DR	InterPro: IPR001507; Endoglin/CD105.				
DR	SMART: SM00241; Zp: 1.				
KW	Cuticle; Signal; Repeat.				
FT	SIGNAL	1	18		POTENTIAL.
FT	CHAIN	19	424		CUTICLIN 1.
FT	DOMAIN	302	323		4 X 4 AA REPEAT OF A-A-P-[AVI].
FT	REPEAT	302	305		1.

FT	REPEAT	307	311	2.
FT <td>REPEAT</td> <td>312</td> <td>315</td> <td>3.</td>	REPEAT	312	315	3.
FT <td>REPEAT</td> <td>320</td> <td>323</td> <td>4.</td>	REPEAT	320	323	4.
FT <td>CONFLICT</td> <td>27</td> <td>27</td> <td>G -> P (IN REF. 1)</td>	CONFLICT	27	27	G -> P (IN REF. 1)
FT <td>CONFLICT</td> <td>29</td> <td>29</td> <td>P -> G (IN REF. 1)</td>	CONFLICT	29	29	P -> G (IN REF. 1)
FT <td>CONFLICT</td> <td>306</td> <td>306</td> <td>A -> G (IN REF. 1)</td>	CONFLICT	306	306	A -> G (IN REF. 1)
FT <td>CONFLICT</td> <td>309</td> <td>309</td> <td>MISSING (IN REF. 1)</td>	CONFLICT	309	309	MISSING (IN REF. 1)
FT <td>CONFLICT</td> <td>389</td> <td>390</td> <td>CL -> SS (IN REF. 1)</td>	CONFLICT	389	390	CL -> SS (IN REF. 1)
FT <td>CONFLICT</td> <td>395</td> <td>396</td> <td>FL -> LP (IN REF. 1)</td>	CONFLICT	395	396	FL -> LP (IN REF. 1)
FT <td>CONFLICT</td> <td>401</td> <td>402</td> <td>FA -> LP (IN REF. 1)</td>	CONFLICT	401	402	FA -> LP (IN REF. 1)
FT <td>CONFLICT</td> <td>401</td> <td>402</td> <td>IG -> MR (IN REF. 1)</td>	CONFLICT	401	402	IG -> MR (IN REF. 1)
EQ <td>SEQUENCE</td> <td>424 AA;</td> <td>45168 MW;</td> <td>2D178A602C1500B5 CRC64;</td>	SEQUENCE	424 AA;	45168 MW;	2D178A602C1500B5 CRC64;

Query Match	70.1%;	Score 1420.5;	DB 1;	Length 424;
Best Local Similarity	67.1%;	Pred. No. 4.5e-119;		
Matches 279;	Conservative 33;	Mismatches 59;	Indels 45;	Gaps 7;

QY	9	CTTATLATS-I-IPDNGVEGEPEIECGTSTTINTNTNAEGVYVYKGLDQSGRRD	67
Db	8	CLALVYLSASAIIPVDNNEGEPEVECGNSTLTVFNTFRPEEGHYVYKGLDYQSGRSD	67
QY	68	GGROYAGSLFDFDSCNARTSLSLPRGILFVTTYVYISHPILFVTKVDRAVVOCFYEMAD	127
Db	68	GGROYAGIELDFDSCNARTSLSLPKGVESTVYISHPDFVTKVDRAVYVIOCFYEMSD	127

```

0Y 128 K1Y5AQIEVSEITLTAQOTQIVMPVCKREILDGPTGQPVQFAITGPVYHKMTCDSETV 187
    ||| |||:::|||||:|||||:|||||:||||| ||| ||| ||| ||| ||| ||| |||
Db 128 K1V5TOIEVSDLTTLTAQOTQIVMPVCKREILDGPGSGQPIQFAITGQDVYHKMTCDSETV 187

```

```
QY 188 DTFCAVHSHSEFVDDGNGDTVELLNADGALDKYLLNNLEPTDLMAGQEAHYKKYADRSQ 247
      |||||  |||||  |||  :|||  :|||  |||||  |||||  |||||  |||||  |||||
Db 188 DTFCAVHSHCTVDDGNGDTVQILNEEGALDKFLNNLEPTDLMAGQEAHYKKYADRSQ 247
```

246	LEFYQQLSILLINPNSLCAVRFQCSSEPGFGAVNITGG-----AA 285
247	:
248	LEFYQQLSITIKDPGSECAVPTCSEPGFGAVNITGGAGAGAAHAAAAPQAGVEEVQAAPVAA 307
Db	

```

308 AABVAAAPVAAAAAAPVPRATIAQLRLLRKKRSEGENEGCITLDVRYEINTLIDMEG--ASP 3655

```

Db 366 SAPEAAALVSE--ESVRRRAISTGICLPIGFASFGLGISTIVATAALSATIFYVARP 419

RESULT 2	
YOH3_CAEEL	
ID	YOH3_CAEEL
YC	0099276.
STANDARD;	PRT; 446 AA.

DE	Hypothetical	50.0	kDa	protein C43C3.3 in chromosome X
DT	16-OCT-2001	(Rel.	40,	Last annotation update)
DT	16-OCT-2001	(Rel.	40,	Last sequence update)
DT	01-NOV-1997	(Rel.	35,	Created)

GN
OS
OC
OC
OC

Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Chromodorea; Rhabditida; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

OR
RN
RP
RC

11
SEQUENCE FROM N.A.
STRAIN-Bristol N2;

RL Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP REVISIONS.
RT

Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collabor
 CC between the Swiss Institute of Bioinformatics and the EMBL/GenBank/DDBJ databases.
 CC

CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no
CC modified and this statement is not removed. Usage by and for comme

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC -----

DR EMBL; 247067; CA87330.2; -.

DR WormPep; C43C3.3; CE23591.

DR InterPro; IPR001507; Endoglin/CD105.

DR SMART; SM00241; ZP; 1.

KW Hypothetical protein; Transmembrane; ATP-binding.

FT TRANSMEM 377 397 POTENTIAL.

FT NP_BIND 122 129 ATP (POTENTIAL).

SO SEQUENCE 446 AA; 49954 MW; ACC469B3B0032DB4 CRC64;

Query Match 5.3%; Score 108; DB 1; Length 446;
Best Local Similarity 22.6%; Pred. No. 0.056;
Matches 77; Conservative 47; Mismatches 128; Indels 88; Gaps 19;

```

QY 31 IEGPSTITNFN-----TRNAEGHYVYKGLYQEGC---RNDEGRQVAGISL 77
      ::  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 34 VDCIADSFYVVLNKSDEVMKRISNPKSQPVVYVYGHKTRHPCGISMKDEKGLTNPMLTI 93

```

02 18 PDS-CVANKIKSLNPRGLFVITTVLSFH-PLFVTKVDRAKRVCCFYMEADKTSQIE 135
 94 PYSECDVLTLDL--PKHRYAETTVVLEDNADLSFGKTRRLNHVFCLYTRNVK-----IR 147

[illegible]

```

      : | : || | : | :: |
Db 206 ISPECTFSDREDISAPDAKITF--VQGGCPNGMND-----IIDPLA 247

227 VPMIMAGOPAHUVVVADNBOI EVDOOTCTMTKEDNSGAPUDNOOGOCSTAYUMCGAAA 286

```

```
D0      : : : | | : : | :  
248 NVNDQIYFSKERTFREGNSQTFVHCVOYLCK--DECSK-TCYKK-----VSD 294  
  
D1      : : : | | : : | :  
287 KPAAAAOILRLKKRSAPENITIDVFRTDINTLFTSDDNOAL 326
```

```

Db      295  SNLTAERLRFRHKRS-----ITDLERRTTSAPYDDNGSL 329
          | : || || | | : | : || | : |

```

ID	GAU1_PSEAE	STANDARD;	PRT;	279	AA.
GAU1_PSEAE					
AC	Q9I291;				

DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE UTP-glucose-1-phosphate uridylyltransferase (EC 2.7.7.9) (UDP-glucos

DE uridylyltransferase) (Uridine diphosphoglucose pyrophosphorylase).
GN GALU OR PA2023.
OS pseudomonas aeruginosa.
PC Bacteria; Proteobacteria; Gammaproteobacteria; Betaproteobacteria; Rhodospirillales; Rhodocyclaceae; Rhodocycla

OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N A

RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Larou M.
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RX MEDLINE=20437337; PubMed=10984043;
RC STRAIN=AWCC 15692 / PAOI;
RX

RA Gaider L.L., Gaulty L., Tolentino E., Westdörck-Waßmann S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
PA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.†

CC -I- FUNCTION: MAY PLAY A ROLE IN STATIONARY PHASE SURVIVAL (BY

CC -1- CATALYTIC ACTIVITY: UTP + alpha-D-glucose 1-phosphate =
CC diphosphate + UDP-glucose.


```
CC -----
CC DR EMBL: U03751; AAB01486.1; -.
CC DR InterPro: IPR001825; NTP_transferase.
CC DR Pfam: PF00483; NTP_transferase; 1.
CC DR TIGRFAMs: TIGR01099; galu; 1.
CC KW Transferase; Kinase; Nucleotidyltransferase.
CC SO SEQUENCE 279 AA; 30939 MW; D8BD1A5A14165FA9 CnC64;
CC
CC Query Match 4.7%; Score 96; DB 1; Length 279;
CC Best Local Similarity 20.8%; Pred. No. 0.35;
CC Matches 49; Conservative 38; Mismatches 93; Indels 56; Gaps 12;
CC
CC QY 19 IPVDN-----GVSEPEIECGPSSITL-NENTRAFECHGYVKLYDQEG-CRNDEGR 70
CC :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
CC Db 28 LPVNRKPLIYGCV--EEALDAGLNEDISIVTGKRALDHEDIS--YELENOIKGTDEK 83
CC QY 71 OVAGISLPFDCSNVARTSLNDRCG---FTTTVVISFHPLEVTAVDRAYRQCIFYMEAD 127
CC :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
CC Db 84 YLVGIIRKLIDECSPSYTQOTOMKGLGHAILGRPLIGDEPPFAVLADL---CYNLESD 139
CC QY 128 KTVSNOIEVSE---ITTAFQTQIYPMPCRYEILDG-----GPTGPQV 167
CC :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
CC Db 140 GVLDGMWKLQKYCKRTITYAAVEVNPTENKKGVINGDDIGDLINRWOMEKPAPEDAPS 199
CC :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
CC QY 168 QPAIIQGPFYHKWTCDSETVDFCAVHSCFVDGNGDTVELNA-----DGCAI 217
CC :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
CC Db 200 NLAIFGRYL-----TPDIFFKIET---EPKGGEIQTTDALLKQAKGCVI 244
CC :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
CC RESULT 5
CC AMD2_XENLA STANDARD; PRT; 875 AA.
CC ID AMD2_XENLA STANDARD; PRT; 875 AA.
CC AC P12890;
CC DT 01-OCT-1989 (Rel. 12, Created)
CC DE 01-OCT-1989 (Rel. 12, Last sequence update)
CC DT 15-JUN-2002 (Rel. 41, Last annotation update)
CC DE Peptidy1-glycine alpha-amadtang monooxygenase II precursor
CC (EC 1.14.17.3) (Peptide C-terminal alpha-amadtang enzyme II) (AE-II).
CC X Xenopus laevis (African clawed frog).
CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
CC OC Xenopodinae; Xenopus.
CC OX NCBI_TaxID=8355;
CC [1]
CC RP SEQUENCE FROM N.A.
CC RC TISSUE=Skin.
CC RX MEDLINE=68134244; PubMed=2829695.
CC RA Onshye K., Kitano K., Wada Y., Fuchimura K., Tanaka S., Mizuno K.,
CC RA Matsuo H.;
CC RT "Cloning of cDNA encoding a new peptide C-terminal alpha-amadtang
CC RT enzyme having a putative membrane-spanning domain from Xenopus laevis
CC RT skin";
CC RL Biochem. Biophys. Res. Commun. 150:1275-1281(1988).
CC -I- FUNCTION: C-terminal alpha-amadtang of peptides. The reaction
CC produces a peptidy1(2-hydroxyglycine) intermediate is unstable and
CC dismutates to glyoxylate and the corresponding desglycine peptide
CC amide.
CC CC -I- CATALYTIC ACTIVITY: Peptidy1glycine + ascorbate + O(2) =
CC peptidy1(2-hydroxyglycine) + dehydroascorbate + H(2)O.
CC CC -I- COFACTOR: COPPER AND ASCORBATE.
CC CC -I- SIMILARITY: BELONGS TO THE COPPER TYPE II, ASCORBATE-DEPENDENT
CC MONOOXYGENASE FAMILY.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC CC -----
CC EMBL: M19032; AAA4967.1; -.
CC PR: A27715; URXIA2.
```

DR HSP: P14925; 1PHM.
 DR InterPro: IPR000323; Cu2_monooxygenase.
 DR InterPro: IPR001258; NHL.
 DR InterPro: IPR00720; Pamonooxygenase.
 DR Pfam: PF01082; Cu2_monooxygen; 1.
 DR Pfam: PF01436; NHL; 4.
 DR Pfam: PF03712; Cu2_monoox_C; 1.
 DR PRINTS: PR00790; PAMONOXGNASE.
 DR PROSITE: PS00084; CU2_MONOOXYGENASE_1; 1.
 DR PROSITE: PS00085; CU2_MONOOXYGENASE_2; 1.
 DR Oxioreductase; Monooxygenase; Copper; Vitamin C; Transmembrane;
 KM Glycoprotein; Signal.
 FT SIGNAL 1 39
 FT CHAIN 40 875
 FT DOMAIN 40 763
 FT TRANSMEM 764 787
 FT DOMAIN 788 875
 FT DOMAIN 358 364
 FT DOMAIN 387 390
 FT DOMAIN 852 856
 FT CAROHD 455 465
 FT CAROHD 662 662
 FT CAROHD 743 743
 SQ SEQUENCE 875 AA; 97084 MW; C07373AF6BF13450 CRC64;

Query Match 4.7%; Score 94.5; DB 1; Length 875;
 Best Local Similarity 21.6%; Pred. No. 2.2;
 Matches 50; Conservative 25; Mismatches 66; Indels 91; Gaps 12;

OY 33 CQPTSTINFRNAPFEGHYVYKGLYDQCGRNDGCRQVAGISLPFSCNVARPTSLNP 92
 DB 530 CQPTDVAADPTGNFVADGY-----CN-SRIMQFSP 560
 OY 93 RGIFV-----TTTVVISFHPFYTKYRAVRYOCFYMEADKTVS 131
 DB 561 NCMFTIMQGETSSNLPFGQFRIPHSLSMISDQGLCVADENG-RICQFIKATGEFVK 619
 OY 133 AOIEVSEI-TTAFQTOIVMPVRCRIEILDGPTGQVQFAITGQPVYHKMTCDSETVDTF 190
 DB 620 -QIKHQEFGREVFAYSYAP-----GG-----VLYAVNGKPYG---DSRPVQGF 659
 OY 191 CAVVHSCFPDQNGDPTVELNMDGALDKYL--LNMLEPPTLMAGQEHVY 240
 DB 660 -----MLNLSNGD-----ILDTFIPARKNFEMPHDIAAGDDGVY 694

RESULT 6

YAY3_SCHPO STANDARD: PRT; 649 AA.
 AC Q10211;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein CAH3.03c in chromosome I.
 GN SPAC4H3.03c.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Williams R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,

RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitz E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skellton J., Simmonds M., Squires R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
 RA Welfens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Filicz C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Leinrach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Medler H., Wambutt R., Purrelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Motlier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurt S.M.,
 RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakowski G.V., Ussery D., Barrell B.G., Nurse P.;
 RA "The genome sequence of Schizosaccharomyces pombe.";
 RL Nature 415:871-880(2002).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (potential).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: 269380; CAA93342.1; -
 DR KM Hypothetical protein; Transmembrane.
 FT TRANSMEM 265 285
 FT TRANSMEM 564 584
 FT TRANSMEM 626 646
 FT SEQUENCE 649 AA; 74488 MW; 0C97C10E603EB3D CRC64;

Query Match 4.6%; Score 93.5; DB 1; Length 649;
 Best Local Similarity 22.1%; Pred. No. 1.8;
 Matches 86; Conservative 52; Mismatches 145; Indels 107; Gaps 21;

OY 25 VEGPEIEEGPSTITINFTRNAF-----EGHYVYVGLYDQCGRNDGCRQVAGISL 77
 DB 36 LDGSEVMACWP-----NFDSPSIFARILDARAGHSITPI-EQTSCK-----QM 78
 OY 78 PFDSCNVARTRSLNPRGIFVTTVVISFH-----PLFVTKVDRA-----YRVOCF 122
 DB 79 YEPSTNIIHTKPYSEKGLV---RLDFFHREWEDYELPLMIRVSCIRGSRKILCEF 135
 OY 123 YMEADKTVSAQIEVSEI-TTAFQTOIVMPVRCRIEILDGPTGQ-----PVQFAI 171
 DB 136 PALDYARQSHETRVSKITENYQAEFVPAAGDPKYLIDCVPSGDLKIDLEIYPAEHLI 195
 OY 172 IGQPVYHKMTCDSEVDFCAVHSCFYVDGNGDPTVELNMDGALDKYLNLNLEXPDL 231
 DB 196 EGGVYISLLEEGGEITF-----IFQEGGFGPVADVTPN-----LVDKLEDSTKR 242
 OY 232 MAGQEAHYKYADRSLQFLYQ---COISITIREPNSECV-RPOCSEPOGFAVKT----- 281
 DB 243 YWRAMIQCCVYTGRRREVQNALUKLITIEPGAVIASPTFSLPEDLGVRWMDYRFT 302
 OY 282 --GGAAPPAALAAQL-----RLIKRSAPPE-NIIDVRTDINTLEISDNO 324
 DB 303 WIRDSAFITVALAQLGFRFAEAVEYMSFTYHVLKRNKKGGINIV-----YSIHGDSQ 354
 OY 325 AL-PYDLRH-RALLQHNQGPVIL--AAVON 350
 DB 355 NLEEVELHLNG--YNSHPVIRIGMAVHH 382

RESULT 7

PYRL_SCHPO STANDARD: PRT; 2244 AA.
 ID PYRL_SCHPO
 AC Q09794;
 DT 01-NOV-1995 (Rel. 32, Created)

Db 1160 -----DTYGYPLVLRPSVYLSCGAMNTYSSQSDLSYLOQANAIKDHVPVITSKIY 1210

QY 244 DRSQLEYQOQISITIKKEPNSCEVRCPOGSEPGGAVKGTGAA---AKPAAAQRLLRK 299

Db 1211 ENAK-----EIEDNVAAREGKMVMHYISEHVENAGVSGDATLVLPQDLAPTIERIYD 1265

QY 300 RBAEPENIIDVKTIDNTLEISDDNQALPYDLRHRALLQHGNGPVIIAANQNGICMSPFEE 359

Db 1266 AAKKIGEAINTGYPYNIQPIADNDEIKVIECNVRA---SRSEPFVSKVI--GVDMTSMAT 1320

QY 360 SMFMG 364

Db 1321 DYING 1325

RESULT 8

NRK1_YEAST

ID NRK1_YEAST STANDARD: PRT: 1080 AA.

AC P38692: 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Serine/threonine-protein kinase NRK1 (EC 2.7.1.37) (N-riich kinase 1).

GN NRK1 OR K1C1 OR YHR102W

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

OX NCBI_TaxID=4932;

RN [1]

RP STRAIN=DC-5;

RA Fukami Y.;

RL Submitted (MAY-1994) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=S288C / AB972;

RX MEDLINE=94378003; PubMed=8091229;

RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J., Du Z., Favell A., Fulton L., Galtung S., Giesel C., Kirsten J., Kucaba T., Hillier L., Jier M., Johnston L., Langston Y., Lacroix P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L., Nhan M., Rifkin L., Riles L., St Peter H., Trevisan E., Vaughan K., Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R., Vaidin M.;

RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome VIII."

RL Science 265:2077-2082(1994).

CC -1- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.

CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; D29980; BAA06250.1; -

DR EMBL; U00059; AAB68860.1; -

DR PIR; S48944; S48944

DR SGD; S0001144; K1C1

DR InterPro; IPR000719; Euk_pkinase.

DR InterPro; IPR002290; Ser_thr_pkinase.

DR Pfam; PF00069; Pkinase.1.

DR ProDom; PD000001; Euk_pkinase.1.

DR SMART; SM00220; S_TKc.1

DR PROSITE; PS00107; PROTEIN_KINASE_ATP.1.

DR PROSITE; PS00108; PROTEIN_KINASE_ST.1.

DR PROSITE; PS50011; PROTEIN_KINASE_DOM.1.

KW Transferase; Serine/threonine-protein kinase; ATP-binding.

FT DOMAIN 23 276 PROTEIN KINASE.

FT NP_BIND 29 37 ATP (BY SIMILARITY).

FT BINDING 52 52 ATP (BY SIMILARITY).

FT ACT_SITE 144 144 BY SIMILARITY.

FT SEQUENCE 1080 AA; 117061 MW; 9989EAF315E0B94 CRC64;

Query Match 4.4%; Score 90; DB 1; Length 1080;

Best Local Similarity 21.5%; Pred. No. 7.4;

Matches 45; Conservative 41; Mismatches 93; Indels 30; Gaps 6;

QY 144 QYQIYPMPCVRCSEIILDGCP-----TGCPVQPAIIGQPYHHWTCDSFEVDFCAVHSC 197

Db 200 KYDINSIGITTYEIAITGNPNPCDVEALRAMQLIKSPRLREDRSYSTLKKEFIAL---C 256

QY 198 EYDQNGDVTVEILNDGCAIDKYLLNNLEYPTDLAAGEAHYKXADRSQLEYQOQISIT 257

Db 257 IDEDPK---ERLSADDLKSKFIRAHKATPTSILKELISRYLLFRDKNKKKIKGSIIP 312

QY 258 IKEPNSCEVRCPOGSEPGGAVKGTGAAPKAAQRLLRKRSAPENIIDVRIIDNTL 317

Db 313 ENEPSKPEAPKPSQ-----NGGDEAKQKSIASNDNEIKRVN---EGDVEMKKWDFSL 362

QY 318 EISD-----DNQALPYDLRHRALLQHN 339

Db 363 SSDDYIENNINLDALAEQNNEMATAQHD 391

RESULT 9

TUD_DROME

ID TUD_DROME STANDARD: PRT: 2515 AA.

AC P25823:

DT 01-MAY-1992 (Rel. 22, Created)

DT 01-MAY-1992 (Rel. 22, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Maternal tudor protein.

GN TUD.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;

OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=92038995; PubMed=1936993;

RA Golubetski G.S., Bardsley A., Tax F., Boswell R.E.;

RT "Tudor, a posterior-group gene of Drosophila melanogaster, encodes a novel protein and an mRNA localized during mid-oogenesis.";

RL Genes Dev. 5:2060-2070(1991).

CC -1- FUNCTION: REQUIRED DURING OOGENESIS FOR THE FORMATION OF PRIMORDIAL GERM CELLS AND FOR NORMAL ABDOMINAL SEGMENTATION.

CC -1- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT THE LIFE CYCLE.

CC -1- MISCELLANEOUS: THE TUD MRNA ACCUMULATES WITHIN THE POSTERIOR REGION OF THE DEVELOPING OOCYTE DURING THE EARLY TO MIDDLE STAGES OF OOGENESIS.

CC -----

CC -1- SIMILARITY: CONTAINS 9 TUDOR DOMAINS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; X62420; CAA44286.1; -

DR PIR; A41519; A41519.

DR HSSP; O16637; IG5V.

DR FlyBase; FBgn0003891; tud.

DR InterPro; IPR001097; Maternal_tudor.

DR InterPro; IPR002999; Tudor.

DR Pfam; PF00567; TUDOR.10.

DR SMART; SM00333; TUDOR.9.

DR PROSITE; PS50304; TUDOR.9.

KW Developmental protein; Repeat.

FT DOMAIN 455 513 TUDOR 1.

```
FT DOMAIN 641 696 TUDOR 2.
FT DOMAIN 1062 1122 TUDOR 3.
FT DOMAIN 1355 1414 TUDOR 4.
FT DOMAIN 1662 1718 TUDOR 5.
FT DOMAIN 1839 1898 TUDOR 6.
FT DOMAIN 2023 2082 TUDOR 7.
FT DOMAIN 2211 2269 TUDOR 8.
FT DOMAIN 2392 2451 TUDOR 9.
SQ SEQUENCE 2515 AA; 285236 MM; 683C100AD308BADA CRC64;

Query Match 4.4%; Score 90; DB 1; Length 2515;
Best Local Similarity 19.1%; Pred. No. 23;
Matches 70; Conservative 51; Mismatches 140; Indels 106; Gaps 16;

QY 16 SYSLPVNNGVEPEIECGPSTI---INFNTRNAFEG--HYVKGLYDOEGCRNDEGG 69
DB 1904 SLQHP-DAVISWPEAEKFAELTGEGLVETQLKPGDHYVIDLDD-----G 1953
QY 70 ROVAGISLPDSCNVAFTSLNPRGIFVTPTVVISFHPLEFVKYDRAVRQCFYMEAD-- 127
DB 1954 ENIIDRLPL--CORKEPKASKESLAVTTKAIT-----THVENTSRITLQFSEKDSL 2004
QY 128 -----KTVSAQIE-----VSEITTAFOQIIVMPYCRY 155
DB 2005 MDICEKLNKSKLQPKTEKAAVDNMCVQFADLEFYSRLLEVLEDDQYKVI----- 2057
QY 156 EILDGQPT-----GQPVQFAITGGPVYHKWTCDSF-----YVDFCAVYHSC 197
DB 2058 -LDYGMTVVDKLYELPQEFLLI-KPYAE--ICSMPSAIFENKKAITLTFPALDSC 2113
QY 198 FVDDGNDGTYEILNADGALDKYLLNNLEPYTDLMAGOEAHVYKYADRSQLFYCOQISIT 257
DB 2114 -----KGVAVAEFVNKASPPVYRLTKDKKSLKLYEHLQKLYVAELKLI 2158
QY 258 IK-EPNSECYAPQCEPQGFQAVKGTGAAPAAQRLILKRSAPENIIDVTDINT 316
DB 2159 QKRENSECEITSYGNSPKSFYQMKHNSADLDLIYKTLQSLKKEKL--KKLIDPTNSNG 2216
QY 317 LEISDDN 323
DB 2217 VCYSQED 2223

RESULT 10
THRC_BUCAI STANDARD; PRT; 429 AA.
AC P57289;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Threonine synthase (EC 4.2.3.1).
GN THRC OR BU192.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OC Bacterioidetes; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=118099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tokyo 1998;
RX MEDLINE=20445173; PubMed=10993077;
RA Shigenobu S., Matanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "genome sequence of the endocellular bacterial symbiont of aphids
BU Buchnera sp. Aps.";
RL Nature 407:81-86(2000).
CC -1- CATALYTIC ACTIVITY: O-phospho-L-homoserine + H(2)O = L-threonine +
phosphate.
CC -1- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC -1- PATHWAY: Threonine biosynthesis.
CC -1- SIMILARITY: BELONGS TO THE SERINE/THREONINE DEHYDRATASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
```

```
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP001118; BAB12909.1; .
DR InterPro; IPR001925; B6_enzyme_beta.
DR InterPro; IPR000634; S/T_dehydrtase.
DR InterPro; IPR004450; Thr_synthase.
DR Pfam; PF00291; PALP; 1.
DR TIGRfams; tigr00260; thrc; 1.
DR PROSITE; PS00165; DEHYDRATASE_SER_THR; 1.
KW Threonine biosynthesis; Lyase; Pyridoxal phosphate; Complete proteome.
FT BINDING 108 108 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ SEQUENCE 429 AA; 48659 MM; F66C73BEF5FAFFC6 CRC64;

Query Match 4.3%; Score 88; DB 1; Length 429;
Best Local Similarity 20.0%; Pred. No. 3.2;
Matches 78; Conservative 57; Mismatches 145; Indels 110; Gaps 17;

QY 14 ALSTIPVNDGVEPEIEC-----GPTSTINFNTR----- 45
DB 79 AFSEKHPLK--IKITKDIHCELPFGPTLAFKDGARFMAQMIILLNKNESVTILRATS 136
QY 46 -----NAFEGH-----VYVKG-----LYDGGCRNDGGRQVAGISL--PFDSCNV 84
DB 137 GDTGAANAANAYGKKNRVRVITLIPKGISLEQELFCTL--GNRIKITSINGSEDDCOK 193
QY 85 ARTSLNPRGIFVTTVVISFHPLEFVKYDRAVRQCFYMEADKTVSAQIEVSEITTAFOQ 144
DB 194 LVKEAFNDKRLKES---IGLSANSINISRLAQICYFAFSLISEQR-----K 241
QY 145 TQIVPMPCREIILDGFTGTPQVQFAITGGPVYHKWTCDS--SEYVDFCAVYHSCFYDDG 202
DB 242 NLVIAVPCGNGNLTAGLSKS---GLPIKSFIACTNANDVPR-----FLNNG 288
QY 203 NGDVEIILNADGALDKYLLNNLEPYTDLMAGOEAHVYKYADRSQLFYCOQISITIKEPN 262
DB 289 TWNPKRTVSTSNMAD-----ISQPNMRYIEELFYKKKMDLKKRGSGSDHTEETL 342
QY 263 SECYAPQCEPQGFQAVKGTGAAPAAQRLILKRSAPENIIDVTD-----INTL 317
DB 343 KLEFPL-----GYSEPHAAIAYRLLRDQKLENRGELFGTAHPAKKNTV 388
QY 318 E-ISDDNALPYDLRHR---ALLQNGQPV 343
DB 389 EKILNKISLPSELQNRIDLPLLSHINIPV 418

RESULT 11
CGEL_CHICK STANDARD; PRT; 407 AA.
AC P49707; Q91032;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE G1/S-specific cyclin E1.
GN CCNE1 OR CCNE.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA Li H., Lahti J.M., Valentine M., Houston J., Kidd V.J.;
RT "The cyclin E gene is apparently essential in avian B cells.";
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: ESSENTIAL FOR THE CONTROL OF THE CELL CYCLE AT THE G1/S
(START) TRANSITION.
CC -1- SUBUNIT: INTERACTS WITH A MEMBER OF THE CDK2/CDK PROTEIN KINASES
TO FORM A SERINE/THREONINE KINASE Holoenzyme complex. THE CYCLIN
CC SUBUNIT IMPARTS SUBSTRATE SPECIFICITY TO THE COMPLEX (BY
```

```

CC -1- SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -1- PTM: PHOSPHORYLATION BY CDK2 TRIGGERS ITS RELEASE FROM CDK2 AND
CC DEGRADATION VIA THE UBIQUITIN PROTEASOME PATHWAY (By similarity).
CC -1- SIMILARITY: BELONGS TO THE CYCLIN FAMILY. CYCLIN E SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U28981; AAA7981.1; -
DR EMBL: U28990; AAA81647.1; -
DR EMBL: U28982; AAA81647.1; JOINED.
DR EMBL: U28983; AAA81647.1; JOINED.
DR EMBL: U28984; AAA81647.1; JOINED.
DR EMBL: U28985; AAA81647.1; JOINED.
DR EMBL: U28986; AAA81647.1; JOINED.
DR EMBL: U28987; AAA81647.1; JOINED.
DR EMBL: U28988; AAA81647.1; JOINED.
DR EMBL: U28989; AAA81647.1; JOINED.
DR HSSP: P30274; IVIN.
DR InterPro: IPR004366; Cyclin.
DR InterPro: IPR004367; Cyclin_Cterm.
DR Pfam: PF00134; cyclin.1.
DR Pfam: PF02984; cyclin_C.1.
DR SMART: SM00385; CYCLIN; 1.
DR PROSITE: PS00292; CYCLINS; 1.
DR CYCLIN: Cell cycle; Cell division; Phosphorylation; Nuclear protein.
KM MOD RES 392 392 PHOSPHORYLATION (BY SIMILARITY).
FT CONFLICT 106 106 G -> GD (IN REF. 1; AAA81647).
FT CONFLICT 343 343 L -> S (IN REF. 1; AAA81647).
SQ SEQUENCE 407 AA: 46739 MW: A1032C7CB0BC1D1A CRC64;

Query Match
Best Local Similarity 4.3%; Score 86.5; DB 1; Length 407;
Matches 65; Conservative 48; Mismatches 101; Indels 103; Gaps 16;

QY 79 FDSQVAFRRS-----LNPRGIVTTTVTSFHPLE-----109
DB 89 FHLNVASTRSSPLILGMANRDVWKNMINKETVYRDLQMRHLLQPKRTILLDW 148
QY 110 ---VTVDRAVRYOCFYEMED-----KTVSAQIEVSEITTTAFQ-TQIYPMRY 152
DB 149 LMEVCEVYKLYR-ETFFLAODEFDFRMAFQOVNVKTLDLIGISSLDFIAKLEIYPPKL 207
QY 153 CRYE-IIDGGPTGQPV--QFAITGQPVYHKWTCDSFVDFCAVV-----194
DB 208 HQFAVYVTDGACTDELLSMELLIMKAL--NMNINPLTVVSMINITYQVAYINLELYVLP 265
QY 195 ---HSCFVDGNGDVEIILNADGALKYLLNNLEYPDLMAQEAHVYKYADRSQI---248
DB 266 QYFQQLFV-----QIALLL--DLCLVDL---IGLEYTYGVLAASALYHFSSELMKVSG 315
QY 249 FYGQCSITIKEPNSRCVRCQSEFGFCAVKTGGAAAPAAAQRLKKSAEPENIT 308
DB 316 YEMCEL-----EECVKMWVPEAM--AIREVGS-----KLKHFRGIAPEDILH 355
QY 309 DVRTDINTLEISDNOQA 325
DB 356 NIOTHTINSIDLDKAQA 372

```

RESULT 12
YAOE_SCHPO STANDARD; PRT; 1260 AA.
AC Q10093; OS9664;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)

```

DE Hypothetical protein C11D3.14c in chromosome I.
GN SPAC11D3.14C.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; Pubmed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Felwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeil C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skellton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grynoprez B.,
RA Welfjens I., Vanstreels E., Rieger M., Schaefer M., Mueller Auer S.,
RA Gabel C., Fuchs M., Fritzc C., Holtzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehnach E., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambert R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Motter S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garçon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Patahkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe."
RL Nature 415:871-880(2002).
CC -1- SIMILARITY: BELONGS TO THE OXOPROLINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z68166; CA92315.1; -
DR InterPro: IPR002821; Hydantoinease_A.
DR InterPro: IPR003692; Hydantoinease_B.
DR Pfam: PF01968; Hydantoinease_A; 1.
DR Pfam: PF02538; Hydantoinease_B; 1.
DR Hypothetical protein.
SQ SEQUENCE 1260 AA: 138784 MW: 44F14F0638224BBB CRC64;

Query Match
Best Local Similarity 4.3%; Score 86.5; DB 1; Length 1260;
Matches 73; Conservative 49; Mismatches 111; Indels 153; Gaps 18;

QY 34 GPTSTINNTNNAEGHYVYK-----GLYDOECGRN---DEGGRQVAGISLPFSDCN 83
DB 856 GITPGSMPSNSKAITYEAGAITFFVYKAGTDEGLQLQLFDEPAK-----YPPDS 907
QY 84 VARTSLN-----PRGIFVTTTVTSFHPLEFYTKYDRAVRYOCFYWEADKTVSA 132
DB 908 GSRTLNDNISDVKAMLSACHGRSRNVEKLYVEGLDIYGR--SMGIGI-----AAA 956
QY 133 QIEVSEITTAFTQTQIVPMPCRYETLIDG-----PTGQPV-QFAITGQPVYHKW 180
DB 957 EKAVRDVYKAFSVQNSQPKLAIDYMDGTPLQLBVKIDPEGTDAVFPEFGTGPVYCNW 1016
QY 181 TCDSTVTFEFCVAVH-----SCFVD-----DGNQDT 206

```

```
Db 1017 --NAPIATYSSVIYLRSLINODIPLNEGKLEIRIPSPCLNPFSETAAYGVNLT 1074
QY 207 VE-----ILNADG-CALDKYLLNNLEYPTDLMAGOEAVHYKYADRSOLFQCOISTITKE 260
Db 1075 SQRITDYLKAFSICASOGCMNLTFCYDGENGE----- 1110
QY 261 PNSECVBPQSEPGFAGVKT--GGAAPK-----AAAAQLRLKRSAPENIIDVRTDI 314
Db 1111 -----GFAMYETIAGAGAGPTWNGTSGVHTMTNTRITDPE-VYERRA-- 1153
QY 315 NTEISDNOALPYDLKRLALQHG 340
Db 1154 -----PVILRRPFLRENSG 1167

RESULT 13
ISCS_METTE STANDARD; PRT; 404 AA.
AC P57795;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable cysteine desulfurase (EC 4.4.1.-) (Nifs protein homolog).
GN ISCS OR NIFS.
OS Methanosarcina thermophila.
OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2210;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=20389605; PubMed=10930739;
RA Borupa B., Ferryb J.G.;
RT "Cysteine biosynthesis in the archaea: Methanosarcina thermophila
RT utilizes O-acetylserine sulfhydrylase.";
RL FEMS Microbiol. Lett. 189:205-210(2000).
CC -1- FUNCTION: CATALYZES THE REMOVAL OF ELEMENTAL SULFUR FROM CYSTEINE
CC TO PRODUCE ALANINE (BY SIMILARITY).
CC -1- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO CLASS-V OF PYRIDOXAL-PHOSPHATE-DEPENDENT
CC AMINOTRANSFERASES. NIFS/ISCS SUBFAMILY.
CC -1- CAUTION: The conserved pyridoxal-binding site Lys at position 216
CC is replaced by a Glu.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF276772; AGO1802.1; -
DR InterPro: IPR000192; AminoTransfV.
DR Pfam: PF00266; aminoTran_5; 1.
DR PROSITE: PS00595; AA_TRANSFER_CLASS_5; FALSE_NEG.
KM Lysase; Pyridoxal phosphate. BY SIMILARITY.
FT ACT SITE 338
FT SEQUENCE 404 AA; 44335 MW; 87B5BA25F87E2A25 CRC64;
SO

Query Match 4.2%; Score 86; DB 1; Length 404;
Best Local Similarity 21.1%; Pred. No. 4.5;
Matches 71; Conservative 37; Mismatches 109; Indels 120; Gaps 16;

QY 123 YMEADKTVSAQIEYSETTAFQOIVPMPQRYEI----- 157
Db 19 YMSASATPVYRKEVYEMLPYMTENFGNPSIYIGKTSKHAIMARKKADALGAENE 78
QY 158 ---LDGPGTG-----QFVOFA-----IIGOPYHKWTCDSFTVDFFCAVHSCFPVDDG 202
Db 79 IYFSGTSEDNAIKIAIRNRKXKHIIITSIEHH-----AVLHICAMIEG 126
QY 203 NGDTVEILAND--GCALDKYLLNNLEYPTDLMA-----GQEAHVY 240
```

```
Db 127 QGFVEVTLVDRKGVNSPDELRLMARDPTLLISIMFANNEIGTIOPIKEIGELAKENOIY 186
QY 241 KYADRSOLFQCOISTITKEPNSECVBPQSE---POEFGA-----YKT-----GGA 286
Db 187 FHTDAVCAI--GHVIDKATKINIDLLSGHEFEGPKCGALYIRKGVKIDPLHGAQE 244
QY 287 KPAANAQLRLKRSAPENI-----IDVRDINTLEISDNOALPYDLRRRL----- 335
Db 245 -----RKRAGENVPGIYGLKATLRLATLAELEENRRL-LKLRRLIEGLLKI 292
QY 336 --LQHGOPYLLAAVONGICMSPFGSKFMGLSTALI 370
Db 293 PRTHLNGHPT--QRLANNVNT---FEYIEGESILL 324

RESULT 14
DSG2_HUMAN STANDARD; PRT; 1117 AA.
AC Q14126;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Desmoglein 2 precursor (HDGc).
GN DSG2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Colon carcinoma;
RX MEDLINE=94192736; PubMed=8143788;
RA Schaefer S., Koch P.J., Franke W.W.;
RT "Identification of the ubiquitous human desmoglein, Dsg2, and the
RT expression catalogue of the desmoglein subfamily of desmosomal
RT cadherins.";
RL Exp. Cell Res. 211:391-399(1994).
RN [2]
RP SEQUENCE OF 777-1117 FROM N.A.
RX MEDLINE=92037656; PubMed=1935985;
RA Koch P.J., Goldschmidt M.D., Walsh M.J., Zimbelmann R., Franke W.W.;
RT "Complete amino acid sequence of the epidermal desmoglein precursor
RT polypeptide and identification of a second type of desmoglein gene.";
RL Eur. J. Cell Biol. 55:200-208(1991).
CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS.
CC INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE
CC FILAMENTS MEDIATING CELL-CELL ADHESION.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).
CC -1- TISSUE SPECIFICITY: ALL OF THE TISSUES TESTED AND CARCINOMAS.
CC -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
CC (POTENTIAL).
CC -----
CC -1- SIMILARITY: CONTAINS 5 CADHERIN DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z26317; CAA8126.1; -
DR HSP; P15116; INCT.
DR Genew; HGNC:3049; DSG2.
DR MIM; 125671; -
DR InterPro: IPR002126; Cadherin.
DR Pfam: PF00028; cadherin; 4.
DR PRINTS: PR00205; CADHERIN.
DR SMART: SM00112; CA_4.
DR PROSITE: PS00232; CADHERIN_1; 3.
DR PROSITE: PS50268; CADHERIN_2; 4.
KW Cell adhesion; Glycoprotein; Transmembrane; Repeat; Signal;
KW Cytoskeleton; Calcium-binding.
```

```
FT SIGNAL 1 23 POTENTIAL.
FT PROPEP 24 48 POTENTIAL.
FT CHAIN 49 1117 DESMOGLEIN 2.
FT DOMAIN 49 608 EXTRACELLULAR (POTENTIAL).
FT TRASMEM 609 633 POTENTIAL.
FT DOMAIN 634 1117 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 49 159 CADHERIN 1.
FT DOMAIN 160 272 CADHERIN 2.
FT DOMAIN 273 387 CADHERIN 3.
FT DOMAIN 388 502 CADHERIN 4.
FT REPEAT 880 911 DESMOGLEIN REPEAT 1.
FT REPEAT 912 941 DESMOGLEIN REPEAT 2.
FT REPEAT 942 967 DESMOGLEIN REPEAT 3.
FT REPEAT 968 991 DESMOGLEIN REPEAT 4.
FT REPEAT 992 1020 DESMOGLEIN REPEAT 5.
FT REPEAT 1021 1050 DESMOGLEIN REPEAT 6.
FT CARBOHYD 111 111 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 181 181 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 308 308 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 461 461 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 513 513 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1117 AA; 122385 MW; 2238897ED70B289 CRC64;
```

Query Match 4.2%; Score 85.5; DB 1; Length 1117;
Best Local Similarity 25.6%; Pred. No. 20;
Matches 52; Conservative 27; Mismatches 81; Indels 43; Gaps 12;

```
OY 15 LLSYIPVDNGVEGEPEI--ECGPTSTITINFNTRNAPEG-----HYVKGLYDQEGGR 64
DB 347 LDFSVYVANKAKAFHKSIRKRYKPTPIPIVKKVKNKKEGHEHSSVSTIVSESMD---R 402
OY 65 NDEGROVAGISLPDSCN---VAFTRSLNPRGIF---VTTVVISFHPLEVTKV-- 113
DB 403 SSKG--QIIGNQAEDEDGTLPAAHAYVYKLEDRDNMISVDSYTSSEIKLKPDESRYVQ 460
OY 114 DDAYRYOCYVMEAD---KTVSAQ--IEVSEITTAFOQVPM-PVCR-----YEILD 159
DB 461 NQTYVYKVAISEDYPRKTIITGTVLINVEDINDNCPITLIEPQVICHDAEYVNTAEDLD 520
OY 160 GSGTGPVQFAIIGQP--VYHKM 180
DB 521 GHPNSGPFSEFVIDKPPGMAEKM 543
```

RESULT 15
HIRA_FUGRU STANDARD; PRT; 1025 AA.

```
ID HIRA_FUGRU STANDARD; PRT; 1025 AA.
AC 042611;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE HIRA protein (Tup1 like enhancer of split protein 1).
GN HIRA OR TUPLE1.
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Neuteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Acanthomorphi; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Takifugu.
CN NCB1_TaxID=31033;
RN [1]
RP SEQUENCE FROM N.A., Pubmed-9524281;
RX MEDLINE-98201624;
RA Llavador R., Estivill X., Scambler P., Pritchard M.;
RT "Isolation and genomic characterization of the TUPLE1/HIRA gene of
the pufferfish Fugu rubripes.";
RL Gene 208:279-283(1998).
CC -1- FUNCTION: Could have a part in mechanisms of transcriptional
CC regulation similar to that played by yeast HIR1 and HIR2 together.
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -1- SIMILARITY: CONTAINS 4 WD REPEATS (TRP-ASP DOMAINS).
CC -1- SIMILARITY: BELONGS TO THE HIR1 FAMILY OF WD-REPEAT PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```
CC EMBL; U94325; AAC60370.1; -.
CC EMBL; U94324; AAC60369.1; -.
CC InterPro: IPR001680; WD40.
DR Pfam: PF00400; WD40; 8.
DR PRINTS: PR00320; GPROTEINBRPT.
DR PRODOM: PD00018; WD40; 3.
DR SMART: SM00320; WD40; 5.
DR PROSITE: PS00678; WD_REPEATS_1; 1.
DR PROSITE: PS00678; WD_REPEATS_2; 3.
DR PROSITE: PS00678; WD_REPEATS_3; 3.
DR PROSITE: PS00678; WD_REPEATS_4; 1.
KW Transcription regulation; Repeat; WD repeat; Nuclear protein.
FT REPEAT 68 98 WD 1.
FT REPEAT 129 159 WD 2.
FT REPEAT 172 202 WD 3.
FT REPEAT 266 313 WD 4.
FT DOMAIN 673 682 POLY-ALA.
FT DOMAIN 685 688 POLY-ALA.
SQ SEQUENCE 1025 AA; 111856 MW; A4212152D75B6A37 CRC64;
```

Query Match 4.2%; Score 85; DB 1; Length 1025;
Best Local Similarity 22.1%; Pred. No. 19;
Matches 44; Conservative 37; Mismatches 80; Indels 38; Gaps 8;

```
OY 200 DDGNDYVEILNADGALDKYLLNNLEYPTDLMAGQEAHVYVYADRSQLFYQCIISITIK 259
DB 604 EDTSSDSEDKM---ATINKNLAFKKRKPELLMDGAEVYERKKKRRP---KDMAASIA 656
OY 260 EPNSECVRCQCEPQGFQVKTGGAAPAAAOURL--LKKR-----SAPENIIDV 310
DB 657 QLTQYTSFAEREPs--RAAAGAGAAAPTAALUKLPTPSIKKFTLOVSDPSVLEV 714
OY 311 RFDINTLEIS-----DDNOALPYDLRHRALLQHNQGPYTLAAVNGICMSPF- 357
DB 715 ENEVSVAAGSRSLQRCSDRGDRDNTLLP-----SSVLTAGSSDVAVASODRLSVFS 769
OY 358 --GFSMEGSLSTALIAAVI 374
DB 770 SCGRRLPALQATPASPAL 788
```

RESULT 16
DYHG_CHLRE STANDARD; PRT; 4485 AA.

```
ID DYHG_CHLRE STANDARD; PRT; 4485 AA.
AC 039575;
DT 01-NOV-1997 (Rel. 35, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Dynein gamma chain, flagellar outer arm.
GN ODA2 OR ODA-2.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
CN NCB1_TaxID=3055;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX STRAIN-1132D;
RA MEDLINE-94274766; Pubmed-7516341;
RA Wilkerson C.G., King S.W., Wiltman G.B.;
RT "Molecular analysis of the gamma heavy chain of Chlamydomonas
flagellar outer-arm dynein.";
RL J. Cell Sci. 107:497-506(1994).
CC -1- FUNCTION: FORCE GENERATING PROTEIN OF EUKARYOTIC CILIA AND
CC FLAGELLA. PRODUCES FORCE TOWARDS THE MINUS ENDS OF MICROTUBULES.
CC DYTENIN HAS ATPASE ACTIVITY.
CC -1- SUBUNIT: CONSISTS OF AT LEAST 3 HEAVY CHAINS (ALPHA, BETA AND
CC GAMMA), 2 INTERMEDIATE CHAINS AND 8 LIGHT CHAINS.
```


CC -1- SIMILARITY: BELONGS TO THE DYNEIN HEAVY CHAIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U15303; AAA50455.1; -.
 CC Interpro: IPR004273; Dynein_heavy.
 CC Interpro: IPR001230; Prenal_site.
 CC Pfam: PF03028; Dynein_heavy; 1.
 CC Motor protein; Microtubules; Dynein; ATP-binding; Flagella;
 CC Coiled coil.
 KW Coiled coil.
 FT DOMAIN 449 469 COILED COIL (POTENTIAL).
 FT DOMAIN 804 838 COILED COIL (POTENTIAL).
 FT DOMAIN 1093 1114 COILED COIL (POTENTIAL).
 FT DOMAIN 1275 1297 COILED COIL (POTENTIAL).
 FT DOMAIN 1699 1727 COILED COIL (POTENTIAL).
 FT DOMAIN 1917 1945 MICROTUBULE-BINDING (POTENTIAL).
 FT DOMAIN 3077 3099 COILED COIL (POTENTIAL).
 FT DOMAIN 3196 3227 COILED COIL (POTENTIAL).
 FT DOMAIN 3265 3343 COILED COIL (POTENTIAL).
 FT DOMAIN 3569 3663 COILED COIL (POTENTIAL).
 FT NP_BIND 1819 1826 ATP (POTENTIAL).
 FT NP_BIND 2099 2106 ATP (POTENTIAL).
 FT NP_BIND 2432 2432 ATP (POTENTIAL).
 FT NP_BIND 2802 2809 ATP (POTENTIAL).
 SQ SEQUENCE 4485 AA; 512836 MW; 974B79328B403677 CRC64;
 Query Match 4.2%; Score 85; DB 1; Length 4485;
 Best Local Similarity 20.8%; Pred. No. 1.4e+02;
 Matches 84; Conservative 51; Mismatches 131; Indels 138; Gaps 24;
 QY 1 MMIRLAF-CCT- LIAL-SYSDPNDGVEGEPEE-CGPT-SITINFN 43
 Db 2771 MKLELVFTDCVTHMTRILAMPGLGLVGVGSGKSLRSATAGTFTYITKYN 2830
 QY 44 TRNAEGHYVYKGLYDEGCR-NDEG-GROVAGISLPEP 80
 Db 2831 VSNLFE-HIKGLYKIAFGGQPVYFLTDAYKDEGFLEYINQILMTGEVAGILTRKD 2887
 QY 81 SCNARFRSLNP-RGIFTVTVYISHPLFVTKV-DRAVOCFTEAKTKYSAQ 133
 Db 2888 QDMT-VNDRPVNKHQAPGLDYYDNLNPF-FLNKRVDNLHVLCF- 2931
 QY 134 IEVSEITTAFOCTIVPMPVCRXYELIDGPTGQPVQFALIGQPVYHKMTCDSEVDFTCAY 193
 Db 2932 -SPVAKAFARRAQFP-GING-C-TIDWFCP- 2959
 QY 194 VHSCEVDGNGDVEIINADGALDKYL-NNLEYPTDLMAQEAHYVYADRS 246
 Db 2960 -GPKRLTSVSGKFDKFTMACPKVEKNQLE-LLMG-NAHVFTAAK 3004
 QY 247 QLF-YOCOSITIKENPSECVRCQSEPOSGAVKTGAAGAARAAQAQRLKKRSERP 304
 Db 3005 EYFKYRYRYVTPKS-YLSFLQYKELIYAKKSTFKELIYOLEVACQKMEP 3056
 QY 305 ENIDVPTDINTL-EISDNOALPVDLRHRLALOHNGOPVILA 346
 Db 3057 -KADYNNKKAELAVKNGTAVSAKKEAALLKQISESTAI 3094
 RESULT 17
 ID CAOC CANFA STANDARD; PRT; 410 AA.
 AC P12637;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Calsequestrin, cardiac muscle isoform precursor.

GN CASQ2.
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 OX NCBI_TaxID=9615;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Heart;
 RX MEDLINE=8243763; PubMed=3379055;
 RA Scott B.T., Sliemers H.K.B., Collins J.H., Nadal-Glunard B.,
 RA Jones L.R.;
 RT "Complete amino acid sequence of canine cardiac calsequestrin deduced
 RT by cDNA cloning.";
 RL J. Biol. Chem. 263:8958-8964(1988).
 RN [2]
 RP PARTIAL SEQUENCE OF 20-71.
 RC TISSUE=Heart;
 RX MEDLINE=88107564; PubMed=3427023;
 RA Shupsky J.R., Ohnishi M., Carpenter M.R., Reithmeier R.A.F.;
 RT "Characterization of cardiac calsequestrin.";
 RL Biochemistry 26:6539-6544(1987).
 RN [3]
 RP PHOSPHORYLATION BY CK2.
 RX MEDLINE=91093153; PubMed=1985907;
 RA Cala S.E., Jones L.R.;
 RT "Phosphorylation of cardiac and skeletal muscle calsequestrin
 RT isoforms by casein kinase II. Demonstration of a cluster of unique
 RT rapidly phosphorylated sites in cardiac calsequestrin.";
 RL J. Biol. Chem. 266:391-398(1991).
 CC -1- FUNCTION: CALSEQUESTRIN IS A HIGH-CAPACITY, MODERATE AFFINITY,
 CC CALCIUM-BINDING PROTEIN AND THUS ACTS AS AN INTERNAL CALCIUM STORE
 CC IN MUSCLE. THE RELEASE OF CALCIUM BOUND TO CALSEQUESTRIN THROUGH
 CC A CALCIUM RELEASE CHANNEL TRIGGERS MUSCLE CONTRACTION. BINDS 40
 CC TO 50 MOLES OF CALCIUM.
 CC -1- SUBCELLULAR LOCATION: THIS ISOFORM OF CALSEQUESTRIN OCCURS IN THE
 CC SARCOPLASMIC RETICULUM/S TERMINAL CISTERNAE LUMINAL SPACES OF
 CC CARDIAC AND SLOW SKELETAL MUSCLE CELLS.
 CC -1- SIMILARITY: BELONGS TO THE CALSEQUESTRIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: J03766; AAA30833.1; -.
 CC PIR: A28071; A28071.
 CC PIR: A39040; A39040.
 CC PIR: C27499; C27499.
 CC HSP: P07221; IABY.
 CC Interpro: IPR001393; Calsequestrin.
 CC Pfam: PF01216; Calsequestrin; 1.
 CC PRINTS: PR00312; CALSEQUESTRIN.
 CC PROSITE: PS00863; CALSEQUESTRIN_1; 1.
 CC PROSITE: PS00864; CALSEQUESTRIN_2; 1.
 CC Muscle; Glycoprotein; Calcium-binding; Signal; Phosphorylation.
 KW SIGNAL 19
 FT CHAIN 20 410 CALSEQUESTRIN, CARDIAC MUSCLE ISOFORM.
 FT DOMAIN 219 254 PRO-RICH.
 FT DOMAIN 221 242 CALCIUM REGULATED HYDROPHOBIC SITE.
 FT DOMAIN 372 410 ASP/GLU-RICH (ACIDIC).
 FT CARBOHYD 335 335 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 395 395 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT MOD_RES 397 397 PHOSPHORYLATION (BY CK2).
 FT MOD_RES 401 401 PHOSPHORYLATION (BY CK2).
 FT MOD_RES 405 405 PHOSPHORYLATION (BY CK2).
 FT CONFLICT 71 71 Q -> E (IN REF. 2).
 SQ SEQUENCE 410 AA; 47416 MW; FCA99A807E7A8B82 CRC64;
 Query Match 4.2%; Score 84.5; DB 1; Length 410;
 Best Local Similarity 20.3%; Pred. No. 6.2;

OY	80	DSCVAVATRLSLNPGIPIPTTVIVISHEHLEPYT-----	111
DB	244	ESSWTGISTGVGRQ---TSTAIVVSGRAYTVSESSSPGTSKESASETTTGISTGSTK	299
OY	112	--KVDRAYRVOCFMEADKTVAISOIEVEITTAFOOTIVPVCRYEILIDGGPTGOPOVF	169
DB	300	SNRITTSRIP--YPETTVATGEQE-TEMTTGCITS.LPPACV-----GPLKEKSP	350
OY	170	AIIOPAPYHKWTC-DSETVD---TFCAVSHSCFEVDGNGDTVELINADCALDKY-----	220
DB	351	GDIWTFANCCKCTCDAELVDCDLKECPSPPKCEBEE---RLVKFKDNTPCEIAACEBPRT	407
OY	221	-LTLNNLEYCPDMLMAGEAHVKKYADRSQLFYOCOSITRIKPENSCVBPPOCSEPFGFAY	279
DB	408	CLENNDY-----EVGASFDPKPKNPCISYSCHN-TGFVAV	441

RESULT_20
TKT_BACHD TKT_BACHD STANDARD; PRT; 666 AA.

ID	OGKAD7:	
DT	16-OCT-2001 (Rel. 40, Created)	
DT	16-OCT-2001 (Rel. 40, Last sequence update)	
DT	16-OCT-2001 (Rel. 40, last annotation update)	
DE	Transketolase (EC 2.2.1.1) (TK).	
GN	TKT OR BH2352.	
OS	Bacillus halodurans.	
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.	
OX	NCBI_TaxID=8665;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=C-125 / JCM 9153;	
RX	MEDLINE=20512582; PubMed=11058132;	
RA	Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N., Fujii F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S., Horikoshi K. ;	
RT	"Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and genomic sequence comparison with Bacillus subtilis."	
RL	Nucleic Acids Res. 28:4317-4331(2000).	
CC	-I- CATALYTIC ACTIVITY. Sedoheptulose 7-phosphate + D-glyceraldehyde 3-phosphate -> D-ribose 5-phosphate + D-xylulose 5-phosphate.	
CC	-I- COFACTOR: THIAMINE PYROPHOSPHATE (By similarity).	
CC	-I- SIMILARITY: BELONGS TO THE TRANSKETOLASE FAMILY.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or send an email to license@lsb.slb.ch).	
CC	-----	
CC	EMBL: AP001515; BABO6071.1; .	
DR	HSSP: P23254; ITRK.	
DR	InterPro: IPRO00360, Transketolase.	
DR	Pfam: PF00456; transketolase; 1.	
DR	Pfam: PF02779; transket_pyr; 1.	
DR	Pfam: PF02780; transketolase_C; 1.	
DR	TIGRFAMs: TIGR00232; tklase_bact; 1.	
DR	PROSITE: PS00801; TRANSKETOLASE_1; 1.	
DR	PROSITE: PS00802; TRANSKETOLASE_2; 1.	
KW	Transferase; Thiamine pyrophosphate; Complete proteome.	
KW	SEQUENCE 666 AA; 72112 MW; EF617PF4FB84D06 CRC64;	

Query Match 4.1%; Score 83; DB 1; Length 666;
Best Local Similarity 2.5%; Pred. No. 16;
Matches 39; Conservative 29; Mismatches 66; Indels 32; Gaps 7;

DB	161	GPTGPVPQVALIIGQPVTHKVTCDSEYVDTFCAVVHSCEFVDGNGDTVIELNADGALDKY	220
DB	116	GPLDGQVAMA-VGNAMERHLAATYNRGNYIMVHYTYTTCGDGLMGVSAEAASLAGH	174

Oy	221	L	-----LNINLEYPDL-----MAGEAHYKRAKDSQFYCOICSIITK	259
Db	175	LKLGRIILLYDSNDISLDLDDLLHHSFSESVEDRPKAGWVHVEDDNNL	-----DEIAKITE	231
Oy	260	EPUSSECVRPOCSEPO	---GFGAVKTSQGAAK---PAAAOVLRLKK	299
Db	232	EAKAD-ERFSLIEVKTITIGFGSPNKGKSVSGAPLGADEVKLTKE	276	
RESULT 21				
ID	PD21_ARATH	STANDARD:	PRT:	820 AA.
AC	P58766:			
DT	15-JUN-2002 (Rel. 41, Created)			
DT	15-JUN-2002 (Rel. 41, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Phospholipase D zeta (EC 3.1.4.4) (AtPLDzeta) (PLD zeta).			
GN	PLDZETA OR AT5G25370 OR F16618.110.			
OS	Arabidopsis thaliana (Mouse-ear cress).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons: core eudicots; Rosidae;			
OC	eucosids II; Brassicales; Brassicaceae; Arabidopsis.			
OX	NCBI_TaxId=3702;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=cv. Columbia;			
RX	MEDLINE=21016721: PubMed=11130714:			
RA	Tabata S., Kaneke T., Nakamura Y., Kotani H., Kato T., Asanizu E.,			
RA	Miyajima N., Sasamoto S., Kimura T., Hosouchi T., Kawashima K.,			
RA	Kohara M., Matsunoto M., Matsuno A., Mureki A., Nakayama S.,			
RA	Nakazaki N., Nario K., Okumura S., Shino S., Takeuchi C., Wada T.,			
RA	Watanabe A., Yamada M., Yasuda M., Sato S., de la Bastide M.,			
RA	Huang E., Spiegel L., Gnoj L., O'Shaughnessy A., Preston R.,			
RA	Huberman K., Murray J., Johnson D., Rohlfing T., Nelson J.,			
RA	Stonick T., Pegin K., Spieth J., Sekhon M., Armstrong J., Becker M.,			
RA	Belter E., Cordum H., Cordes M., Courtney L., Courtney W., Dante M.,			
RA	Du H., Edwards J., Fryman J., Haakensen B., Lamar E., Latreille P.,			
RA	Leonard S., Meyer R., Mulvaney E., Ozerky P., Riley A., Stromatt C.,			
RA	Wagner-McPherson C., Wollam A., Yeakum B., Bell M., Dedhia N.,			
RA	Parnell L., Shah R., Rodriguez M., Hoon See L., Vill D., Baker J.,			
RA	Kirchoff K., Toth K., King L., Bahret A., Miller B., Marra M.,			
RA	Mattiesen R., McComble W.R., Wilson R.K., Murphy G., Bancroft I.,			
RA	Volckaert G., Wambut R., Duesterhoelt A., Stiekema W., Pohl T.,			
RA	Enlian K.-D., Terryn N., Hartley N., Bent E., Johnson S.,			
RA	Langham S.-A., McCullagh B., Robben J., Grymopreuz B., Zimmermann W.,			
RA	Ranspenger U., Medler H., Balke K., Medler E., Peters S.,			
RA	van Staveren M., Dikse W., Woolman P., Klein Lankhorst R.,			
RA	Weitzneger T., Bothe G., Rose M., Hauf J., Bernaisser S., Hempel S.,			
RA	Feldpausch M., Lambert S., Villarreal R., Gietlen J., Ardlies W.,			
RA	Bents O., Lemcke K., Kolesov G., Mayer K.F.X., Rudd S., Schoof H.,			
RA	Schneider C., Zaccaria P., Mewes H.-W., Beyan M., Franz P.,			
RT	"Sequence and analysis of chromosome 5 of the plant Arabidopsis			
RT	thaliana."			
RL	Nature 408:826-826(2000).			
CC	-1 FUNCTION: Hydrolyzes glycerol-phospholipids at the terminal			
CC	phosphodiesteric bond.			
CC	-1 CATALYTIC ACTIVITY: A phosphatidylcholine + H(2)O = choline + a			
CC	phosphatidate.			
CC	-1 COPACITOR: Calcium (By similarity).			
CC	-1 SUBCELLULAR LOCATION: Cytoplasmic and membrane-associated (By			
CC	similarity).			
CC	-1 DOMAIN: C2 domain is a calcium-binding fold, and the binding			
CC	promotes the protein association with membranes. A lower affinity			
CC	toward calcium can be anticipated for PLD alpha due to the absence			
CC	of two potential calcium ligands.			
CC	-1 SIMILARITY: Belongs to the phospholipase D family. C2-PLD			
CC	subfamily.			
CC	-1 SIMILARITY: CONTAINS 1 C2 DOMAIN.			
CC	-1 SIMILARITY: CONTAINS 2 PLD PHOSPHODIESTERASE DOMAINS.			
CC	-----			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			

CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL: AC006258; ; NOT_ANNOTATED.CDS.
DR	PROSITE: P550004; C2_DOMAIN_2; FALSE_NEG.
DR	PROSITE: P550035; PLD; 2.
KM	Hydrolase: Lipid degradation; Calcium: Repeat; Multigene family.
CC	-----
FT	DOMAIN 1 117 C2_DOMAIN.
FT	DOMAIN 334 371 PLD PHOSPHODIESTERASE 1.
FT	DOMAIN 662 689 PLD PHOSPHODIESTERASE 2.
FT	ACT_SITE 339 339 POTENTIAL.
FT	ACT_SITE 341 341 POTENTIAL.
FT	ACT_SITE 346 346 POTENTIAL.
FT	ACT_SITE 667 667 POTENTIAL.
FT	ACT_SITE 669 669 POTENTIAL.
FT	ACT_SITE 674 674 POTENTIAL.
SO	SEQUENCE 820 AA; 93362 MW; PFB9247276BCEB7B CRC64;
Query Match	4.1%; Score 83; DB 1; Length 820;
Best Local Similarity	22.6%; Pred. No. 22;
Matches 72; Conservative 48; Mismatches 102; Indels 96; Caps 22;	
OY	54 VKGLVDEGCRNDEGROVAGISLDFSCONVART-RSLNPGIFVTTTVVLSFPHLEVTX 112
Db	46 IKRLLD--SCTSLFSGHLXATIDL--DSRAKRTMRMRHPWL-----QSFVYTAHS 94
OY	113 VDRAYRVOCFYMEADKTVSAQ-----IEVSEITTAFOFOTIVPMPCRY-EILDGSPGTQ 165
Db	95 ISKI-----IFVKEDEPVASALIGRAVLPTVEVING-----QPIDRMIDILD-ENRR 141
OY	166 PQGFALIGQPVYH---KKTCDSEYVDTCGAVYHSGFVDGNGDVEYLIN-ADGCAL---- 217
Db	142 PIQ-----GSGKLHVAKCFKTHVTDVNMNKGIIILPSF---NGVNAVYFNOREGKVTLLXQ 193
OY	218 DRYLLNLEFYPTDLAAGDEAHVYKVA-----DRSOLFQOCQISIR-----IKEP 261
Db	194 DAHVLYN--EYRPVTLGGQV-IYKHHRCWEELFDALINPAKLLIITYAGSVATDVTLYVRP 250
OY	262 NSECVRPQCSSEPFQGFCAVKTGGAAAPRAAAQRLRLKRSAPENIIDVRTDINTLEISD 321
Db	251 KR--TRPG-----GDLKLG-----ELTKKKADEVENTVL-----MLWMD 281
OY	322 DNGALPVDLRHRLALQHN 339
Db	282 DRTSHEVERKRDGLMATHD 299
RESULT 22	
PAPA_HUMAN	STANDARD; PRT; 1627 AA.
ID	PAPA_HUMAN
AC	Q13219; Q08371; Q9UDK7;
DDT	16-OCT-2001 (Rel. 40, Created)
DDT	16-OCT-2001 (Rel. 40, Last sequence update)
DT	15-JUN-2002 (Rel. 41, Last annotation update)
DE	Pregnancy-associated plasma protein-A precursor (EC 3.4.24.-) (PAPP-A)
DE	(Insulin-like growth factor-dependent IGF binding protein-4 protease)
DE	(IGF-dependent IGFBP-4 protease) (IGFBP-4ase).
GN	PAPA.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX	NCBI_TaxID=9606;
RA	SEQUENCE FROM N.A., AND INDUCTION.
RA	TISSUE=Placenta;
RA	MEDLINE=96203921; Pubmed=8620868;
RA	Haaning J., Oxvig C., Overgaard M.T., Ebbesen P., Kristensen T.,
RA	Solltrup-Jensen L.;
RT	"Complete cDNA sequence of the preproform of human pregnancy-
RT	associated plasma protein-A. Evidence for expression in the brain and
RT	induction by CAMP".


```

Query Match          4.1%; Score 83; DB 1; Length 1627;
Best Local Similarity 19.9%; Pred. No. 54;
Matches 65; Conservative 51; Mismatches 119; Indels 92; Gaps 15;

QY 105 FHPLFTKDRARVYOCFMEADKYSAOIEVSEITTAFOQTIVPMPCVREYELDGG--P 162
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 624 FHSFETFPYNN-----FMSYAD-----DDCTDSTPNQVARNHCYLDLVYQWGP 668

QY 163 TCGPVOFAITGCPVYHKWTKDSEVDFCAVYHSCFVDGNGDTVEILNA-----DGCAL 217
   : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 669 SKRPAPVALAPQVIGH--TTDSVITLEMFPPI-----DGHFERRLGSAHCUCLEGRTL 719

QY 218 KYLLNNLE-----YPTDLMAGOBAAHYKYADRSOLFQCCQISITTEPNS---ECVRPQ 269
   : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 720 VQYASNASSPMDCSPSGHMSPREAECHPDVEQ-----PCKSVRTWSPNSAVNPHTPPPA 774

QY 270 GSEPOG-----FGAVKGTGAAMAKPPAAAOQLRL-----KKRSAEPEN 306
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 775 CREPOQCIYELFELPLVPEESLTIWTFVSTMDSS--GAVNDIKLAVSGKNISLGPON 832

QY 307 II-----DVRTDINTLEISDNOALPYDLRHRALLQHNGQFVILAAVONGICM-- 354
   : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 833 VECADVPLTRIMDVGEVYGIOIYTLDEHLEID-----AAMLTSTADTPLCLQC 881

QY 355 SPFGESMFMGLSIALIAVITITISKEF 381
   : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 882 KPLKRYKVVADPPLQMDVASTLHLNRF 908

RESULT 23
SYN_BACHD
ID SYN_BACHD STANDARD; PRT; 430 AA.
AC 09KCT8;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Asparaginyl-tRNA synthetase (EC 6.1.1.22) (Asparagine--tRNA ligase)
DE (AenRS).
GN ASNS OR BH1696.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_Taxid=86665;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fujii F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
CC -1- CATALYTIC ACTIVITY: ATP + L-asparagine + tRNA(Asn) = AMP +
CC diphosphate + L-asparaginyl-tRNA(Asn).
CC -1- SUBUNIT: Homodimer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL: AP001512; BAB05415.1; -
DR HSP, Q52428.188A.
DR InterPro: IPR002106; AATRNA_ligaseII.
DR InterPro: IPR004522; ASNS.
DR InterPro: IPR004364; tRNA-synt_2.
DR InterPro: IPR002312; tRNA-synt_asp.
DR InterPro: IPR004365; tRNA_antL.

```

```

DR Pfam: PF00152; tRNA-synt_2; 1.
DR Pfam: PF01336; tRNA_antL; 1.
DR PRINTS: PR01042; TRNASYNTHASP.
DR TIGRPFAMs: TIGR00457; asns; 1.
DR PROSITE: PS00862; AA-TRNA_LIGASE_II; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; ligase; ATP-binding;
KW complete proteome.
SQ SEQUENCE 430 AA; 49113 MW; E5C6D1D1868E902DB CRC64;

Query Match          4.0%; Score 81.5; DB 1; Length 430;
Best Local Similarity 18.9%; Pred. No. 12;
Matches 65; Conservative 59; Mismatches 123; Indels 97; Gaps 17;

QY 12 LIALSYSTIPVDNG--VEGEPEIECG--PISITINFTRNAFEHGVYKGLYDQGGCNDG 68
   : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 136 IIRATYEFHENGFPKVDPLILGSAPEGETELFPHK-----YFDDAFLSOG 184

QY 69 GROVAGISLPDSCNAVTRSLNPRGIEVTTTVISFHPLEVTKYDRAVR--VQCFYMEA 126
   : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 185 QLYMEAAALAFGR-----VFSFGTFPRAEKSKEIRHLIEFMWIEP 224

QY 127 DKT-----VSAQIEVSEITTAFOQTIVPMPCVREYELDGGTGPQVFAITGCPVYHKWCD 183
   : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 225 EMAYVEFESESLTQENYVAIVQSV-LKHCAIEL-----KTLGR-----D 263

QY 184 SETVDTFCAVHSCFVDGNGDTVEILNADGALKDKYLLNNLEYPDLMAGOBAAHYKYA 243
   : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 264 TSVLES-----IQAPPRISYDDAIKFLHEKG-----FDDEIEMDDDFGAPHEFAIEHF 312

QY 244 DRSOLFQCCQISI-----TIKEPNSCVRPOCE---PQGFCAVKTGGAAMAKPPAAAOQLRL 296
   : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 313 DKPVEITTHYPTSLKPFYMEPPDRDDV-VLCADLIAPEGYRI-TGGS---QRISDYDL 366

QY 297 LKRSAPENIIDVVRTDINTLEISDNOALPYDLRHRALLQHNG 340
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 367 LKKRIIEH-----DLSDIAVAWYIDLKKGYSVPHSG 397

RESULT 24
DPOE_HUMAN
ID DPOE_HUMAN STANDARD; PRT; 2286 AA.
AC 007864; Q13533;
DT 01-OCT-1994 (Rel. 30, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE DNA polymerase epsilon, catalytic subunit A (EC 2.7.7.7) (DNA
DE polymerase II subunit A).
GN POLE OR POLE1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 48-51; 876-886 AND 1338-1344.
RX TISSUE=T-cell;
RX MEDLINE=93252906; PubMed=8486689;
RA Kesti T., Frontali H., Syvaola J.E.;
RT "Molecular cloning of the cDNA for the catalytic subunit of human DNA
RT polymerase epsilon.";
RL J. Biol. Chem. 268:10238-10245(1993).
RN [2]
RN REVISIONS.
RA Syvaola J.E.;
RN submitted (JUN-1998) to the EMBL/Genbank/DBJ databases.
RP [3]
RP SEQUENCE FROM N.A.
RA Asahara H., Goldsmith J.S., Lee E., Lim S.;
RL Submitted (FEB-1996) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: PARTICIPATES IN DNA REPAIR AND IN CHROMOSOMAL DNA
CC REPLICATION.
CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + [DNA](n).
CC -1- SUBUNIT: CONSISTS OF TWO SUBUNITS (258 kDa AND 55 kDa).

```

```

CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- DOMAIN: THE DNA POLYMERASE ACTIVITY DOMAIN RESIDES IN THE
CC N-TERMINAL HALF OF THE PROTEIN, WHILE THE C-TERMINUS IS NECESSARY
CC FOR COMPLEXING SUBUNITS B AND C. THE C-TERMINUS MAY ALSO REGULATE
CC THE CATALYTIC ACTIVITIES OF THE ENZYME.
CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY. HIGH
CC SIMILARITY WITH FUNGAL DNA POLYMERASE II.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; S60080; AAA15448.1; ALT_SEQ.
DR EMBL; L09561; AAC19148.1; -.
DR EMBL; U49356; AAA90924.1; -.
DR PIR; A46692; A46692.
DR GenBank; HGNC:9177; POLE.
DR MIT; 174762; -.
DR InterPro; IPR002064; DNA_POL_B.
DR Pfam; PF00136; DNA_POL_B; 1.
DR SMART; SM00486; POLBc; 1.
DR Transferrase; DNA-directed DNA polymerase; DNA replication; DNA repair;
KM DNA-binding; Zinc-finger; Nuclear protein.
FT ZN_FING 2158 2190 C4-TYPE (POTENTIAL).
FT ZN_FING 2221 2238 C4-TYPE (POTENTIAL).
FT SQDQNC 2286 AA; 261530 MW; A216FELAD137DEC CRC64;
SQ
Query Match 4.0%; Score 81.5; DB 1; Length 2286;
Best Local Similarity 22.5%; Pred. No. 1.2e+02;
Matches 42; Conservative 26; Mismatches 70; Indels 49; Gaps 8;
QY 153 CRVEILDGPTGPVQPAIIGQPY-----HKMTCDSEVDTFCVAVHSCFVD 201
DB 1076 CRV--IIRKKEGSPVTERAIPDAIFQAEPYRKHFRLKRWLSSLODF----- 1122
QY 202 GNGDVTVEILNAD-----GCALDKYL-----LNLEYPTDLAMG-----OEAHYK 241
DB 1123 ---DIRAILMDVYIERLGSIOKIIITIPALQOVKNPVPVVKHPMDLHKLEKNVYK 1179
QY 242 YADRSOLF-YOCQISITIKENSCVBPQCEPGFAGVTKGGAAPAAQAQRLKKR 300
DB 1180 QKTSLEFTEBGRQVYMAESBDSPPRSAPDMDFGLVLPDPA-PVTVKRRRVLMES 1238
QY 301 SAEPEPI 307
DB 1239 QEESQDL 1245

```

```

RT for operon."
RL Mol. Microbiol. 11:1169-1179(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426517; PubMed=9278503;
RA Blattner F.R., Plunkett G., Iii, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Bada T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map."
RL DNA Res. 3:137-155(1996).
RN [4]
RP SEQUENCE OF 767-848 FROM N.A.
RC STRAIN-K12;
RX MEDLINE=94134696; PubMed=8302830;
RA Deguchi C., Kakeda M., Yamada H., Mizuno T.;
RT "An analogue of the DnaI molecular chaperone in Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:1054-1058(1994).
CC -1- FUNCTION: Reduces trimethylamine-N-oxide (TMAO) into
CC trimethylamine; an anaerobic reaction coupled to energy-yielding
CC reactions.
CC -1- CATALYTIC ACTIVITY: NADH + trimethylamine-N-oxide = NAD(+) +
CC trimethylamine + H(2)O.
CC -1- COFACTOR: Molybdenum (Molybdopterin) (By similarity).
CC -1- SUBUNIT: Interacts with the N-terminal domain of tofC.
CC -1- SUBCELLULAR LOCATION: Periplasmic.
CC -1- SIMILARITY: BELONGS TO THE PROKARYOTIC MOLYBDOPTERIN-CONTAINING
CC OXIDOREDUCTASE FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; X73888; CAAS2095.1; -.
DR EMBL; AE000201; AAC74082.1; -.
DR EMBL; D90736; BAA36139.1; -.
DR EMBL; D90737; BAA35764.1; -.
DR EMBL; D16500; -. NOT_ANNOTATED_CDS.
DR PIR; S34222; S34222.
DR HSP; O87948; ITMO.
DR EcGene; EG11814; torA.
DR InterPro; IPR001467; Prok_Mboxred.
DR Pfam; PF00384; molybdopterin; 1.
DR Pfam; PF01568; Molybdop_binding; 1.
DR TIGRfams; TIGR00509; bisco_fam; 1.
DR PROSITE; PS00551; MOLYBDOPTERIN_PROK_1; FALSE_NEG.
DR PROSITE; PS00490; MOLYBDOPTERIN_PROK_2; 1.
DR PROSITE; PS00932; MOLYBDOPTERIN_PROK_3; 1.
KW Oxidoreductase; NAD; Molybdenum; periplasmic; Signal;
KW Complete proteome.
FT SIGNAL 1 39
FT CHAIN 1 40 TRIMETHYLAMINE-N-OXIDE REDUCTASE 1.
FT CONFLICT 173 173 L -> R (IN REF. 1).
FT CONFLICT 176 176 A -> R (IN REF. 1).
FT CONFLICT 256 256 A -> R (IN REF. 1).

```

```

FT CONFLICT 258 258 V -> S (IN REF. 1).
FT CONFLICT 281 281 R -> G (IN REF. 1).
FT CONFLICT 325 325 Q -> E (IN REF. 1).
FT CONFLICT 348 348 T -> S (IN REF. 1).
FT CONFLICT 503 504 KL -> NV (IN REF. 1).
FT CONFLICT 713 714 QQ -> HE (IN REF. 1).
FT CONFLICT 751 751 P -> M (IN REF. 1).
FT CONFLICT 781 781 P -> L (IN REF. 1 AND 3).
SQ SEQUENCE 848 AA; 94456 MW; 59DDACB00B1843E7 CRC64;

Query Match
Best Local Similarity 25.28; Pred. No. 34;
Matches 58; Conservative 22; Mismatches 72; Indels 78; Gaps 15;

QY 42 FNTNAAE-----GHVYKGLVD-----OEGCRNDEGGQVAGISLP-FDS 81
Db 558 FEARNDEDIFELCRFRNREAEFEGIDEMGKMRINDEGQVQKGC---RGVHLPADD 613
QY 82 CNVARTSLNPRGIFVTTTVYSF-HP-LVYTKVDRAYRVQCFYMEAD---KTVSAQIE 135
Db 614 -----FMNNKEYVEFDHPOMFV-----RHQAFREDPLELGTGSLIE 652
QY 136 VSEITTAFCQT---QIYPMPCRYEILDDGGTGPVQFAIIGQPVY----HKVTCSET 186
Db 653 IYKTIADMYDDCGCHPMFPEKERSHGSGSO---KYPHLQSVHPDFRLHSQLCSET 710
QY 187 VDFCAVY--HSCFVDD-----GNQDYVELINADGAL-----DKY 220
Db 711 LRQOYVAGKEPVFINPDASRGIRNGDVYRVFNANGVLAGVNSDXY 760

RESULT 26
CARB_XYLFA STANDARD; PRT; 1080 AA.
ID CARB_XYLFA
AC Q9PECL1;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-
GN carbonyl-phosphate synthase ammonia chain).
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=9a5c;
RC MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvarenga R., Alves L.M.C., Araya J.E., Bala J.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Brliones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carter H.,
RA Colauto N.B., Colombo C., Costa P.F., Costa M.C.R., Costa-Neto C.H.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
RA Pacinencini A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Gantier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hohenseil J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Kleger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Montelero-Vicorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A., Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Pexoto B.R., Pereira G.A.G., Pereira H.A., Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E., Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A., Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,

```

```

RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Mieldanis J., Seubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL Nature 406:151-159(2000).
CC -1- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
CC phosphate + L-glutamate + carbamoyl phosphate.
CC -1- COFACTOR: Binds three manganese ions (By similarity).
CC -1- PATHWAY: Arginine biosynthesis.
CC -1- SUBUNIT: Composed of two chains; the small (or glutamine) chain
CC promotes the hydrolysis of glutamine to ammonia, which is used by
CC the large (or ammonia) chain to synthesize carbamoyl phosphate (By
CC similarity). BELONGS TO THE CARB FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE003946; AAF83917.1; -.
CC HSSP; P00968; 1A9X.
DR InterPro; IPR005483; CPase_L.
DR InterPro; IPR005479; CPase_L_D2.
DR InterPro; IPR005480; CPase_L_D3.
DR InterPro; IPR005481; CPase_L_N.
DR InterPro; IPR004362; MGS_Like.
DR Pfam; PF00289; CPase_L_chain; 2.
DR Pfam; PF02786; CPase_L_D2; 2.
DR Pfam; PF02787; CPase_L_D3; 1.
DR Pfam; PF02142; MGS; 1.
DR PROSITE; PRO00098; CPASE.
DR PROSITE; PS00866; CPASE_1; 1.
DR PROSITE; PS00867; CPASE_2; 2.
KW Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
KW ATP-binding; Manganese; Complete proteome.
FT DOMAIN 1 403 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 404 554 OLIGOMERIZATION DOMAIN.
FT DOMAIN 555 942 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 943 1080 ALLOSTERIC DOMAIN.
FT REPEAT 1 554
FT REPEAT 555 1080
FT NP_BIND 153 210 ATP (POTENTIAL).
FT NP_BIND 303 354 ATP (POTENTIAL).
FT METAL 285 285 MANGANESE 1 (BY SIMILARITY).
FT METAL 299 299 MANGANESE 1 AND 2 (BY SIMILARITY).
FT METAL 301 301 MANGANESE 2 (BY SIMILARITY).
FT METAL 301 301 MANGANESE 3 (BY SIMILARITY).
FT METAL 830 830 MANGANESE 3 (BY SIMILARITY).
FT METAL 847 847 MANGANESE 3 (BY SIMILARITY).
SQ SEQUENCE 1080 AA; 117383 MW; D8CED71DD42328B7 CRC64;

Query Match
Best Local Similarity 4.08; Score 81; DB 1; Length 1080;
Matches 80; Conservative 20.68; Pred. No. 47;
Matches 80; Mismatches 143; Indels 112; Gaps 20;

QY 50 GHVYKGLVD--EGCRNDE-----GROVAGISLPDSCNVAARTSLNPRGIFVTT 100
Db 540 GTATLYXSTYEECEAAPSDDRRKIMILGGPNRIGCGIEFYCVAAALALREDF---ET 596
QY 101 VVISFHPLEF--TKVDRAYRVQCFYMEADKTVSAQIEVSEI-----TT 141
Db 597 IMVNCNPEVSTVDYDTSRL---YFE-PLTLEDVLEIVEVHPGVYVYGGQTPDLAK 652
QY 142 AFQIYPMPCRYEILDDGGTGPVQFAI---IGQPVYHKVTCDESETVDT----- 189
Db 653 ALEANGVPVIGTSPESIDLEDRERFQKLVQGLRQRP-----NCTARFAEALVAREI 708
QY 190 -FCAVVNSCFDDGNG-----DYEIILNADGALDKYLLNNLEYPDTDLMA 233
Db 709 GYPLVVRPSYVLGGRAMEIYVGEADLARVYRDAAVKSNDSPVLLDRFLDNAVEVDVITA 768

```


FT PEPTIDE 749 954 C3C FRAGMENT.
FT PEPTIDE 955 1303 C3D FRAGMENT.
FT PEPTIDE 955 1001 C3E FRAGMENT.
FT PEPTIDE 1002 1303 C3D FRAGMENT.
FT PEPTIDE 1304 1320 C3E FRAGMENT.
FT SITE 748 749 CLEAVAGE (BY C3 CONVERTASE).
FT SITE 1303 1304 CLEAVAGE (BY FACTOR I).
FT SITE 1320 1321 CLEAVAGE (BY FACTOR I).
FT DOMAIN 693 728 ANAPHYLATOXIN-LIKE.
FT DISULFID 559 816 INTERCHAIN (BY SIMILARITY).
FT DISULFID 626 661 BY SIMILARITY.
FT DISULFID 693 720 BY SIMILARITY.
FT DISULFID 694 727 BY SIMILARITY.
FT DISULFID 707 728 BY SIMILARITY.
FT DISULFID 873 1513 BY SIMILARITY.
FT DISULFID 1101 1158 BY SIMILARITY.
FT DISULFID 1358 1489 BY SIMILARITY.
FT DISULFID 1389 1458 BY SIMILARITY.
FT DISULFID 1506 1511 BY SIMILARITY.
FT DISULFID 1518 1590 BY SIMILARITY.
FT DISULFID 1537 1661 BY SIMILARITY.
FT DISULFID 1637 1646 BY SIMILARITY.
FT CARBOHYD 939 939 N-LINKED (GLCNAC. . .).
FT CARBOHYD 1617 1617 N-LINKED (GLCNAC. . .).
FT THIOLEST 1010 1013 BY SIMILARITY.
SQ SEQUENCE 1663 AA; 186482 MW; DE5346CC769BEA19 CRC64;

Query Match 4.0%; Score 81; DB 1; Length 1663;
Best Local Similarity 24.2%; Pred. No. 84;
Matches 65; Conservative 39; Mismatches 97; Indels 68; Gaps 17;

OY 117 YRQCCYMEADKTV-SAQLEVSE-TTTAFQTOIVPMPCRYELLDGPGPGQPVQRIIGQ 174
DB 205 WKIRAYEHAHPKQIFSAEEVKEVLPSEVREPTETFFY-ID-DPNG--LEVSIIAK 259
OY 175 PYHKTCDESEY-----DFECAVVHS--CFVDDGNDTVEILNADGCAIDKYLNN 224
DB 260 ELKGNVDGTAFYIFGVQDGDKKISLASHLTRVIEDGVGDV-----LTRKVLME 310
OY 225 LEVPT--DLMAQGEAHV-----XYADRSQI--FYQCOISITIKPEPSECV 266
DB 311 GVRPSNADALVGLKSLVSVTVILHSGSDVVEARSGIPITVSPYQIHFKTRP-----FF 365
OY 267 RQ-----CSEPGFGFAVK-----TGGAAPKPA---AAQOLRLKRSAPENITDV 310
DB 366 KPAPEMDLWVFYTPDGPSPASKVLVVTGCSNKAALTQDDGVAKLSINTENSRQPLT-ITV 424
OY 311 RTDINTLEISDDNQALPVDLRHRLALLQHN 339
DB 425 RTKKDTL--PESRQATKTMEAHPISTMHN 451

RESULT 28
RLAO_CABEL
ID RLAO_CABEL STANDARD; PRT; 311 AA.
AC O93572;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 60S acidic ribosomal protein p0.
GN RPA-0 OR P25H2.10.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Bristol N2;
RA Wilkinson J.;
RA Submitted (NOV-1996) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-15.
RC STRAIN-Bristol N2;

RX MEDLINE-97295239; PubMed-9150941;
RA Bini L., Heid H., Liberatori S., Geier G., Pallini V., Zwilling R.;
RT "Two-dimensional gel electrophoresis of Caenorhabditis elegans
homogenates and identification of protein spots by microsequencing.";
RL Electrophoresis 18:557-562(1997).
CC -I- FUNCTION: RIBOSOMAL PROTEIN P0 IS THE FUNCTIONAL EQUIVALENT
CC OF E. COLI PROTEIN L10.
CC -I- SUBUNIT: P0 FORMS A PENTAMERIC COMPLEX BY INTERACTION WITH
CC DIMERS OF P1 AND P2.
CC -I- SIMILARITY: BELONGS TO THE L10P FAMILY OF RIBOSOMAL PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; 279754; CAB02098.1; -
DR Siena-2DPAGE; O93572; -
DR WormPep; P25H2.10; CE09655.
DR InterPro; IPR001813; 60S-ribosomal.
DR InterPro; IPR001790; Ribosomal_L10.
DR Pfam; PF00428; 60S-ribosomal; 1.
DR Pfam; PF00465; Ribosomal_L10; 1.
KW Ribosomal protein; Phosphorylation.
FT INTR MET 0
SQ SEQUENCE 311 AA; 33642 MW; 2FA9A35CD24DE0F4 CRC64;

Query Match 4.0%; Score 80.5; DB 1; Length 311;
Best Local Similarity 22.6%; Pred. No. 9.8;
Matches 49; Conservative 24; Mismatches 69; Indels 75; Gaps 10;

OY 138 EITFAFQTOIVPMPCRY--YEIL-----DGGPNS-----OPVOFAIIGQPV 176
DB 132 EKTSFPAQLOIPFKIRGTEILNDVHLKEDKYGASALLNMGVTPFSYGLVVRQV 191
OY 177 YHKTCDESEYVDFCAVHSCFVDDGNDTVEILNADGCAIDKYLNN-----LE 226
DB 192 Y-----DDGTLTYPTFLDWTPELRLKRFSGVRNVAASVLAVN 229
OY 227 YPTDLMAQGEAHVYKADNSQLFYQCOISITIKPEPSECVRQCSFGFAVKTGAAA 286
DB 230 YPT---LASVAHSLANGLOMLGVAAVTVDSERE--AEFIKAFIADPSKFAAAPAAAAA 284
OY 287 KPAAPAAQRLRLKRSAPENITDVRTDINTLEISDN 323
DB 285 -PAAAP-----AAKKEEPK-----ESDD 304

RESULT 29
SYN_STAM
ID SYN_STAM STANDARD; PRT; 430 AA.
AC O99035;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Asparaginyl-tRNA synthetase (EC 6.1.1.22) (Asparagine--tRNA ligase)
GN (ASNRS). SAV1454 OR SA1287.
OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and
OS Staphylococcus aureus (strain N315).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=158878; 158879;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Mu50 / ATCC 700699, and N315;
RA MEDLINE-21311952; PubMed-11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iian J.-O., Ito T.,
RA Kanamori M., Matsumaru H., Matuyama A., Murakami H., Hoshiyama A.,
RA Mizutani Y., Takahashi N.K., Sawano T., Inoue K.-I., Kaito C.,

```

0C Spermatophyta: Magnoliophyta: eudicotyledons; core eudicots; Rosidae;
0C eucotylids 1; Fabales; Fabaceae; Papilionoideae; Vitaceae; Pisum.
0X NCBI_X12388;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=cv. Birte;
RA MEDLINE=90147555; Pubmed=2515855;
RX Ealing P.M., Casey R.;
RT "The cDNA cloning of a pea (Pisum sativum) seed lipoxygenase.
RT Sequence comparisons of the two major pea seed lipoxygenase
RT isoforms".
RL Biochem. J. 264:929-932(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Birte;
RA Casey R.;
RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PLANT LIPOXYGENASE MAY BE INVOLVED IN A NUMBER OF
CC DIVERSE ASPECTS OF PLANT PHYSIOLOGY INCLUDING GROWTH AND
CC DEVELOPMENT, PEST RESISTANCE, AND SENESCENCE OR RESPONSES TO
CC WOUNDING. IT CATALYZES THE HYDROPEROXYDATION OF LIPIDS, CONTAINING
CC A CIS, CIS-1,4-PENTADIENE STRUCTURE.
CC -1- CATALYTIC ACTIVITY: linoleate + O(2) = (9Z,11E)-(13S)-13-
CC hydroperoxyoctadeca-9,11-dienoate.
CC -1- Cofactor: IRON.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE LIPOXYGENASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL, X17061; CAA34906.1; -.
DR EMBL, X78580; CAA5318.1; -.
DR PIR, S07075; SOA7075.
DR HSSP, P08170; 2SBL.
DR InterPro: IPR000907; lipoxygenase.
DR InterPro: IPR001024; lipoxygenase_LH2.
DR Pfam: PF00305; lipoxygenase_1.
DR Pfam: PF01477; PLAT; 1.
DR PRINTS; PR00087; LIPOXYGENASE.
DR SMART; SM00308; LH2; 1.
DR PROSITE; PS00081; LIPOXYGENASE_2; 1.
DR PROSITE; PS00711; LIPOXYGENASE_1; 1.
KW Oxidoreductase; Dioxxygenase; Iron; Multigene family.
FT METAL 524 524
FT METAL 529 529
FT METAL 529 529
FT METAL 716 716
FT METAL 864 864
FT METAL 864 864
FT CONFLICT 333 333
FT CONFLICT 362 362
FT CONFLICT 578 578
FT CONFLICT 578 578
SQ SEQUENCE 864 AA; 97133 MM; 2919AF5FAF272CDF CRC64;
Query Match 4.0%; Score 80.5; DB 1; Length 864;
Best Local Similarity 18.3%; Pred. No. 39;
Matches 48; Conservative 29; Mismatches 76; Indels 109; Gaps 9;
OY 36 TSITINFTNRNFAFEHHVYVKGGLDDEGCRNDEGGRQVAGISLIPDSQNWARTRSLNPRGI 95
DB 298 SAEFDNF--TPNEFDSFOVNRNLE-----GKIKPLD----- 328
OY 96 FVTTTIVISFHEPLFTVKYDRAVRVOCFYMEADKTVSAOIEVSEITTAFOQIVMPVCRY 155
DB 329 -----VISITSLPLPVVK-----EIFRTGGEV-----LKTPTPHVIRV 361
OY 156 EILDGPPGQPVQPAIIIGCPVYHKKTQSEYVDITFCAYVHSCV----- 199
DB 362 S-----KSAMMTDEEFAREMIAGVNPDMCMIQLOEFPKSMIDPAE 401

```

QY 200 --DDGGDVEILNADGCAIDKYLNNLEXPND-----IMAGOEAHVYKADRSOLF 249
 DB 402 YGDHRSKISVDVNLGDGCTIDALSGRLFDYHDFIFPFRINETSARAKATPTILF 461
 QY 250 YO-----COISTIKEPNSE 264
 DB 462 LKENGTAKVALELSLPHPDGD 483

RESULT 31
 TORA_ECO57
 ID TORA_ECO57 STANDARD: PRT: 848 AA.
 AC P58360:
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Trimethylamine-N-oxide reductase 1 precursor (EC 1.6.6.9) (TMAO reductase 1) (Trimethylamine oxidase 1).
 GN TORA OR Z1415 OR ECS1152.
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
 OC NCBI_TaxID=83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D., Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A., Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L., Grobbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamocitis K., Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C., Welch R.A., Blattner F.R.;
 RA "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
 RT Nature 409:529-533(2001).
 RL [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K., Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T., Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T., Kubara S., Shiba T., Hattori M., Shinagawa H.;
 RA "Complete genome sequence of enterohaemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12.";
 RT DNA Res. 8:11-22(2001).
 RL [1]
 CC -1- FUNCTION: Reduces trimethylamine-N-oxide (TMAO) into trimethylamine; an anaerobic reaction coupled to energy-yielding reactions (By similarity).
 CC -1- CATALYTIC ACTIVITY: NADH + trimethylamine-N-oxide = NAD(+) + trimethylamine + H(2)O.
 CC -1- COFACTOR: Molybdenum (Molybdopterin) (By similarity).
 CC -1- SUBUNIT: Interacts with the N-terminal domain of torC (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE PROKARYOTIC MOLYBDOPTERIN-CONTAINING OXIDOREDUCTASE FAMILY.
 CC -1- SUBCELLULAR LOCATION: Periplasmic (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE PROKARYOTIC MOLYBDOPTERIN-CONTAINING OXIDOREDUCTASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF005293; AAC5554.1; -
 DR EMBL: AP002554; BAB34575.1; -
 DR InterPro: IPR001467; ProX-Mboxred.
 DR Pfam: PF00384; Molybdopterin; 1.
 DR Pfam: PF01568; Molybdopterin; 1.
 DR TIGRfams: TIGR00509; disc_fam; 1.

DR PROSITE: PS00551; MOLYBDOPTERIN_PROK_1; FALSE_NEG.
 DR PROSITE: PS00490; MOLYBDOPTERIN_PROK_2; 1.
 DR PROSITE: PS00932; MOLYBDOPTERIN_PROK_3; 1.
 DR Oxidoreductase; NAD; Molybdenum; Periplasmic; Signal;
 KW Complete proteome.
 FT SIGNAL 1 39 BY SIMILARITY.
 FT CHAIN 40 848 TRIMETHYLAMINE-N-OXIDE REDUCTASE 1.
 FT SEQUENCE 848 AA; 9446 MW; ABFFD02ED178932F CRC64;
 SQ

Query Match 3.9%; Score 80; DB 1; Length 848;
 Best Local Similarity 25.2%; Pred. No. 42;
 Matches 58; Conservative 20; Mismatches 74; Indels 78; Gaps 15;

QY 42 ENTRNAFE-----GHVYVKGLYD-----QECGRNDEGROVAGISLP-FDS 81
 DB 558 FEARNDPDFRELCKRFNEBEAFTEGLDGMGLKRWGCGYQGGK-----RGVHLPAFED 613
 QY 82 CNAVARTSLNPGICFYTTTVVISF-HP-LFVTKVDRAVYQCFYMEAD---KTVSAQIE 135
 DB 614 -----FWNNKEVEFFDPQMFV-----RHQAFFREDPDLEPLGTSGLTIE 652
 QY 136 VSEITTAFTQ-----OIVPVCVREYELDGGPGPGVQFALIGOPY-----HKWCDSET 186
 DB 653 IYKTIIDMNVDCCGHPMFEKIERSHGSPQSQYPLHL--QSVHPDFRLHSQCESET 710
 QY 187 VDFPCAIV--HSCFYVD-----GNGDFVEILNADGCAI-----DKY 220
 DB 711 LMQQTVAGKEVFPINPDASARGIRNGDVAVFANRQVYLAGAVVSDRY 760

RESULT 32
 AND_HUMAN
 ID AND_HUMAN STANDARD: PRT: 974 AA.
 AC P13021:
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Peptidyl-glycine alpha-amidating monooxygenase precursor (EC 1.14.17.3) (PAM).
 GN PAM.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Thyroid carcinoma;
 RX MEDLINE=90290494; PubMed=2357221;
 RA Glauder J., Ragg H., Rauch J., Engels J.W.;
 RA "Human peptidylglycine alpha-amidating monooxygenase: cDNA, cloning and functional expression of a truncated form in COS cells.";
 RT Biochem. Biophys. Res. Commun. 169:551-558(1990).
 RL [2]
 RN SUPFAM
 RX MEDLINE=94193805; PubMed=8144680;
 RA Yun H.Y., Keutmann H.T., Bipper B.A.;
 RA "Alternative splicing governs sulfation of tyrosine or oligosaccharide on peptidylglycine alpha-amidating monooxygenase.";
 RT J. Biol. Chem. 269:10946-10955(1994).
 RL [1]
 CC -1- FUNCTION: C-terminal alpha-amidation of peptides. The reaction produces a peptidyl(2-hydroxyglycine) intermediate is unstable and dismutates to glyoxylate and the corresponding desglycine peptide amide.
 CC -1- CATALYTIC ACTIVITY: Peptidylglycine + ascorbate + O(2) = peptidyl(2-hydroxyglycine) + dehydroascorbate + H(2)O.
 CC -1- COFACTOR: COPPER AND ASCORBATE.
 CC -1- SUBCELLULAR LOCATION: Secretory granules.
 CC -1- SIMILARITY: BELONGS TO THE COPPER TYPE II, ASCORBATE-DEPENDENT MONOOXYGENASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its

ID YEM2_YEAST STANDARD; PRT; 1753 AA.
AC P32634;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE Hypothetical 195.4 kDa protein in RPS26B-GLC7 intergenic region.
GN YER132C OR SYCP-OR50.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SOURCE FROM N.A.
RA Mulligan J.T., Dietrich F.S., Hennessey K.M., Sehl P., Komp C.,
RA Wei Y., Taylor P., Nakahara K., Roberts D., Davis R.W.;
RL Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA SRRAIN-S288C / AB972;
RA Dietrich F.S., Mulligan J.T., Hennessey K.M., Allen E., Araujo R.,
RA Aviles E., Berno A., Brennan T., Carpenter J., Chen E., Cherry J.M.,
RA Chung E., Duncan M., Guzman E., Hartzell G., Hunkle-Smith S.,
RA Hyman R., Kayser A., Komp C., Lashkari D., Law H., Lin D.,
RA Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P., Oh C.,
RA Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T., Smith V.,
RA Taylor P., Wei Y., Yelton M., Botstein D., Davis R.W.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO S.POMBE RAL2.
CC -1- SIMILARITY: TO YEAST MD53.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb.ch).
CC -----
DR EMBL: J018916; AAC03230.1; -
DR FIR: S30855; S30855.
DR SGI: S0000934; PMD1.
DR InterPro: IPR001798; Kelch.
DR Pfam: PF01344; Kelch; 2.
KW Hypothetical protein.
SQ
SEQUENCE 1753 AA; E42529986C62500B CRC64;
Query Match 3.9%; Score 79.5; DB 1; Length 1753;
Best Local Similarity 21.6%; Pred. No. 1.2e+02;
Matches 63; Conservative 37; Mismatches 94; Indels 97; Gaps 15;
QY 12 LIALSY--SIPVNGVE-----GEPEIEG-----PISITINENTRNA--FECHV 52
DB 1002 LLSNYSGSDIPYASIOEYGMNNGRDEEGDQDYGICISPNIRIFSTINININGNF 1061
QY 53 YVAGLYDOEGCRNDEGGRVAGISLPFDSQNVARTRS-----INPRGIFY--TT 99
DB 1062 KEEFEFSKSYINNEKSRISYSNPE---SVSTNSNNNAITLEPLLPRLPWST 1118
QY 100 TVVISHPLFVTKVDRAYRVQCFMEADKTVSAQIEVSEITTAFOQIYPMPCREYILD 159
DB 1119 ASVRAAEFFEYF-----AQINGKMLLAVTLDDLIMAKI-YEI-- 1155
QY 160 GGTGTGPVQPAITGQPVYHKTKTDSFTVDFCAVHVSF-----VDDGNG----- 204
DB 1156 -----PILEYLLELVKTIISKEGGLSVTCALNLTLQOKVSRYNCEGKIRKQLDS 1209
QY 205 -----DTVEI-----LNADGALDKYLLNNLEVEPTDLMAGEAHVYKADRS 246
DB 1210 SESYQDTLEIKRSLANIDNGYDSTYLLRNTS-----MAQSH--YTDSS 1251

ID PYR1_YEAST STANDARD; PRT; 2214 AA.
AC P07259;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE URA1 protein [includes: glutamine-dependent carbamoyl-phosphate
GN synthase (EC 6.3.5.5); Aspartate carbamoyltransferase (EC 2.1.3.2)].
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA SRRAIN-ATCC 28583 / F1100;
RX MEDLINE-89378778; PubMed-2570735;
RA Souciet J.-L., Nagy M., le Gouar M., Lacroite F., Potier S.;
RT "Organization of the yeast URA2 gene: identification of a defective
RT dihydroorotase-like domain in the multifunctional carbamoylphosphate
RT synthetase-aspartate transcarbamylase complex.";
RL Gene 79:59-70(1989).
RN [2]
RP SEQUENCE OF 1-510 FROM N.A.
RX MEDLINE-87286375; PubMed-3039294;
RA Souciet J.-L., Potier S., Hubert J.-C., Lacroite F.;
RT "Nucleotide sequence of the pyrimidine specific carbamoyl phosphate
RT synthetase, a part of the yeast multifunctional protein encoded by
RT the URA2 gene.";
RL Mol. Gen. Genet. 207:314-319(1987).
RN [3]
RP SEQUENCE OF 1268-2214 FROM N.A.
RX MEDLINE-89252578; PubMed-2498313;
RA Nagy M., le Gouar M., Potier S., Souciet J.-L., Herve G.;
RT "The primary structure of the aspartate transcarbamylase region of
RT the URA2 gene product in Saccharomyces cerevisiae. Features involved
RT in activity and nuclear localization.";
RL J. Biol. Chem. 264:8366-8374(1989).
RN [4]
RP SEQUENCE OF 1-276 FROM N.A.
RX SRRAIN-S288C / FY1679;
RC MEDLINE-96408771; PubMed-8948101;
RA Czepluch C., Kordes E., Pujol A., Jauniaux J.-C.;
RT "Sequencing analysis of a 40.2 kb fragment of yeast chromosome X
RT reveals 19 open reading frames including URA2 (5' end), TRK1, PBS2,
RT SPT10, GCD14, RPE1, PHO86, NCA3, ASF1, COT7, GZE3, two tRNA genes,
RT three remnant delta elements and a Ty4 transposon.";
RL Yeast 12:1471-1474(1996).
RN [5]
RP SEQUENCE OF 175-2214 FROM N.A.
RX SRRAIN-S288C / FY1679;
RC MEDLINE-96408771; PubMed-8813765;
RA Katsoulou C., Tzeremla M., Tavernarakis N., Alexandraki D.;
RT "Sequence analysis of a 40.7 kb segment from the left arm of yeast
RT chromosome X reveals 14 known genes and 13 new open reading frames
RT including homologues of genes clustered on the right arm of
RT chromosome XI.";
RL Yeast 12:787-797(1996).
CC -1- FUNCTION: THIS PROTEIN IS A "FUSION" PROTEIN ENCODING THREE
CC ENZYMIC ACTIVITIES OF THE PYRIMIDINE PATHWAY (GATASE, CPSASE,
CC AND ATCASE).
CC -1- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
CC phosphate + glutamate + carbamoyl phosphate.
CC -1- CATALYTIC ACTIVITY: Carbamoyl phosphate + L-aspartate = phosphate
CC + N-carbamoyl-L-aspartate.
CC -1- PATHWAY: Pyrimidine biosynthesis; first step.
CC -1- PATHWAY: Pyrimidine biosynthesis; second step.
CC -1- MISCELLANEOUS: GATASE (GLUTAMINE AMIDOTRANSFERASE) AND CPSASE
CC (CARBAMOYL PHOSPHATE SYNTHASE) FORM TOGETHER THE
CC GLUTAMINE-DEPENDENT CPSASE (GD-CPSASE) (EC 6.3.5.5).
CC -1- MISCELLANEOUS: IN EUKARYOTES EC 6.3.5.5 IS SYNTHESIZED BY TWO
CC PATHWAY-SPECIFIC (ARGININE AND PYRIMIDINE) UNDER SEPARATE CONTROL.
CC -1- SIMILARITY: THE CPSASE DOMAIN IS SIMILAR TO OTHER CPSASES.
CC -1- SIMILARITY: IN THE CENTRAL SECTION; BELONGS TO THE DHOASE FAMILY.

```

FT CONFLICT 1592 1592 V -> G (IN REF. 1 AND 3).
FT CONFLICT 1595 1595 S -> A (IN REF. 1 AND 3).
FT CONFLICT 1937 1937 A -> R (IN REF. 3).
FT CONFLICT 1937 1937 T -> I (IN REF. 3).
FT CONFLICT 1997 1997 H -> L (IN REF. 3).
FT CONFLICT 2039 2039 T -> L (IN REF. 3).
FT CONFLICT 2158 2165 KIILHAHE -> VRSWTHOOK (IN REF. 3).
SQ SEQUENCE 2214 AA; 2245124 MW; 4C4A5304DAECAD21 CR664;

Query Match 3.9%; Score 79.5; DB 1; Length 2214;
Best Local Similarity 18.2%; Pred. No. 1,7e+02;
Matches 70; Conservative 52; Mismatches 150; Indels 113; Gaps 16;

QY 47 AFEEHHVYVKKGLYDDGCGNDESGROY-----AGISLPFSCNVAFRSLNPKRIFPTT 99
DB 959 AYTLYMTLYTNADSHDSLFDHGVNLGSGYRIGSSVEFPWCATVAVTRLANNI---K 1015
QY 100 TVVLSFPHLEFTKVDRAVYOCFYMADKTVSAQIEVESEITTAPOTO-----IYMPVCR 154
DB 1016 TIMVNTND-----ETVSTYDEADRLYFETINLERVLDIYEIENSSGVYSM----- 1062
QY 155 YEIIDGGFTGPVOFAI-----IGOPYVHKWTC 182
DB 1063 -----GGGTSSNNIAMTLHRENVKILGTSPMDISAEKNKFRSMLDQGVDPAPAKELTSS 1117
QY 183 DSEFTVDFTC-----AVVHSCFVDDGNG-DIVEIILNADGALDKYLLNNLEYPIDLMAGQ 235
DB 1118 MDE-AESAEKVGKVPVLRPSYVLGSAAMNTVYSKN-----DLESYLNQAAVEVSRD----Y 1168
QY 236 EAHYVKKYADRKQLYQCOISITTIKEPNESECRCPOSEPGQGVAVTGTGAANKPAAQAOLR 295
DB 1169 PVVTKTKYIENK-----EIEMDAVARNSELVHVVSEVENAGVHSGDGYTL----- 1214
QY 296 LAKRSASEPNI-----IDVPTDINTLEISDDNQALPVDLRHRRLALDHNGOP 342
DB 1215 IVPPODLAPPEYVDRIYVATATIGKALKTKTGTGYNIOTFIKADNEIKYIECNVRA---SRSEF 1271
QY 343 VILAAYONGICMSPFGESFMGLSI 367
DB 1272 FISKVW--GVNLIJELATKAIMGDLPL 1294

RESULT 36
VF61_MENJA STANDARD; PRT; 553 AA.
AC 058956;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MJ1561.
ID GN MJ1561.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RA RN
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Kellum G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Sultberg A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reisch C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Goudek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrman J.M., Nguyen D.,
RA Uutterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Eiserich C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
CC Science 273:1058-1073(1996).
CC
CC -1- SIMILARITY: TO M.JANNSCHII MJ0795 AND MJ1506
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

Db 608 -----GIDTV-IADMLPEAKVDYIQR--LRDQGHV-AMVGGINDGPALACADL 653

Db 654 GLAMRGSTVVAIGAADLLV 673

```

RESULT 38
CACH_MOUSE
ID CACH_MOUSE STANDARD; PRT; 827 AA.
AC 09100;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cadherin-17 precursor (Liver-intestine-cadherin) (LI-cadherin) (BILL-
  cadherin) (P130).
GN CDH17.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Angres B., Kim U., Tauber R.;
RT "Li-cadherin gene expression during intestinal development.";
RN Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.; SEQUENCE OF 26-33: 52-58: 74-81: 117-123 AND
  490-509. TISSUE SPECIFICITY, AND DEVELOPMENTAL STAGE.
RC STRAIN-BALB/c; TISSUE-Fetal Liver;
RX MEDLINE=20469471; Pubmed=10906147;
  Ohnishi K., Shimizu T., Karasuyama H., Melchers F.;
RT "The identification of a nonclassical cadherin expressed during B cell
  development and its interaction with surrogate light chain.";
RN J. Biol. Chem. 275:31134-31144(2000).
CC -1- FUNCTION: CADHERINS ARE CALCITONIN DEPENDENT CELL ADHESION PROTEINS.
  THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
  MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
  SORTING OF HETEROGENEOUS CELL TYPES. LI-CADHERIN MAY HAVE A ROLE
  IN THE MORPHOLOGICAL ORGANIZATION OF LIVER AND INTESTINE.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (potential).
CC -1- TISSUE SPECIFICITY: HIGHEST EXPRESSION IS FOUND IN INTESTINE WITH
  LOWER EXPRESSION IN SPLEEN, BONE MARROW, LUNG AND TESTIS. NO
  EXPRESSION DETECTED IN LIVER, KIDNEY, HEART, BRAIN OR SKELETAL
  MUSCLE. EXPRESSED IN PRECURSOR B-CELLS AND MYELOID CELLS.
CC -1- DEVELOPMENTAL STAGE: EXPRESSION INCREASES IN PRO- AND PRE-B-1
  CELLS, DECREASES IN LARGE AND SMALL PRE-B-II CELLS, AND INCREASES
  AGAIN IN IMMATURE AND MATURE B-CELLS.
CC -1- SIMILARITY: CONTAINS 7 CADHERIN DOMAINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation-
  CC the European Bioinformatics Institute. There are no restrictions on its
  CC use by non-profit institutions as long as its content is in no way
  CC modified and this statement is not removed. Usage by and for commercial
  CC entities requires a license agreement (See http://www.isb.ch/announce/
  CC or send an email to license@isb-sib.ch).
-----
DR EMBL: AF177669; AAD51125.1; -
DR EMBL: D87912; BAB03264.1; -
DR HSSP: P15116; INCI.
DR MGD: MGI:1095414; Cdh17.
DR InterPro: IPR002126; Cadherin.
DR Pfam: PF00028; cadherin; 7.
DR PRINTS: PR00205; CADHERIN.
DR SMART: SM00112; CA: 5
DR PROSITE: PS00232; CADHERIN_1; 3.
DR PROSITE: PS00268; CADHERIN_2; 5.
KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
  Signal; Transport.
FT SIGNAL 1 25
FT CHAIN 26 827 CADHERIN-17.
FT DOMAIN 26 766 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 767 807 POTENTIAL.
FT DOMAIN 808 827 CYTOPLASMIC (POTENTIAL).

```

```

FT DOMAIN 29 127 CADHERIN 1.
FT DOMAIN 128 243 CADHERIN 2.
FT DOMAIN 244 339 CADHERIN 3.
FT DOMAIN 340 448 CADHERIN 4.
FT DOMAIN 449 565 CADHERIN 5.
FT DOMAIN 566 666 CADHERIN 6.
FT DOMAIN 667 776 CADHERIN 7.
FT CARBOHYD 148 148 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 249 249 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 418 418 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 545 545 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 586 586 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 721 721 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 827 AA; 91645 MW; CDCECEA5A76E2B58 CRC64;

Query Match 3.9%; Score 79; DB 1; Length 827;
Best Local Similarity 20.2%; Pred. No. 50;
Matches 65; Conservative 36; Mismatches 93; Indels 128; Gaps 16;

QY 130 VSAQIEVSEITTAFTQ-----QIVPVCRRYILDGPTGQ-----PVQ 168
  ||||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 2 VSAQLHFACLTLYLTCGSGEGKFSGLPKM---TFSIFGQPSQVIFQKTNPAVT 58

QY 169 FAIIQGP-----VYHKTCDSFT-----VTFCAVHSCFYDGNQDTV 207
  : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 59 FELGEGDGIPIKIKDGLYHTRALDRETRAHHILQALADS-----HGATVDGPVPTI 113

QY 208 EILNADGALDKYLLNLEPYPTDLMAQEAHV-----YKYAD-----RSOL 248
  : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 114 EVKVI-----NDNRTFLQSKYESVRSNRGRPFMYVNAITDLDPAIPNQL 162

QY 249 FYQCQISITTEPNSECVPCQSPQGFGA-VKTGAAPKPAALQRLKRS----- 301
  ||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 163 FYQIVIDL-----PQINDVMYFQIDSKTGAISLTFEGSGQELDPVKNPSYNIUVS 211

QY 302 -----APEN-----IIDVRTDINTLEISDNQALPDLRHRALLQNGRPVLAIVQ- 349
  : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 212 VKDMGSGENSESPTTYVDVISIRENIWKAPE-----PVEIRENSTDPH--PIKITOVW 263

QY 350 -----NGICMSPFGFSM 361
  : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 264 NDPEAQYSLVKKELSPPEPSI 285

RESULT 39
TYCB_BACBR
ID TYCB_BACBR STANDARD; PRT; 3587 AA.
AC 030408;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrocidine synthetase II [includes: Arg-dependent proline adenylase
  (Proa) (Proline activase); Arg-dependent phenylalanine adenylase
  (Phea) (Phenylalanine activase); Arg-dependent D-phenylalanine
  adenylase (D-Phea) (D-phenylalanine activase); Phenylalanine racemase
  [Arg-hydrolyzing] (EC 5.1.1.11)].
GN TYCB.
OS Bacillus brevis.
OC Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Brevibacillus.
OX NCBI_TaxID=1393;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 8185;
RX MEDLINE=98012987; Pubmed=9352938;
RA Mootz H.D., Marahiel M.A.;
RT "The tyrocidine biosynthesis operon of Bacillus brevis: Complete
  nucleotide sequence and biochemical characterization of functional
  internal adenylation domains.";
RT J. Bacteriol. 179:6843-6850(1997).
CC -1- FUNCTION: ACTIVATES THE SECOND TO FOURTH AMINO ACIDS IN TYROCIDINE
  (IN TYROCIDINE A, PRO, PHE, AND D-PHE) AND EPIMERIZES THE LAST
  ONE.

```



```

Db 79 LFAKMEDTDEV---GVALDVLSPNLISTSMLGKVKYNLSLSDT-----ATGLIQYA 128
QY 244 -----DRSOLFYOQOISITIKENSECVRPOCSEPOGFGAVKTGGAAPK---AAAQLR 295
Db 129 LDGCVNVTQVFEVD---TVGMPEYQARLQOSFPGIEVTVKAKADALYPVSAASICAK 183
QY 296 LKKRSAPENITDVRTDINTLEISDNOALPYDLRHRALLQHNQOPVILAAVQNGICMS 355
Db 184 VARDQAVKKWQVEKLODLDT---DYSGGYPNDDPKTKAWLKEHVEPV----- 227
QY 356 PFGFSMEFGLS 366
Db 228 -FGFPQFVRES 237

```

Search completed: March 24, 2003, 08:49:56
 Job time : 34 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 24, 2003, 08:50:43 ; Search time 35 Seconds
(without alignments)
2278.292 Million cell updates/sec

Title: US-10-054-562a-4

Perfect score: 2026
Sequence: 1 MIMRIACFTLIALSYIP.....ALIAVITIRISRFENOKA 387

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPREMBL_21:*

- 1: sp.archaea:*
- 2: sp.bacteria:*
- 3: sp.fungi:*
- 4: sp.human:*
- 5: sp.invertebrate:*
- 6: sp.mammal:*
- 7: sp.mhc:*
- 8: sp.organelle:*
- 9: sp.phage:*
- 10: sp.plant:*
- 11: sp.rodent:*
- 12: sp.virus:*
- 13: sp.vertebrate:*
- 14: sp.unclassified:*
- 15: sp.virus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1653.5	81.6	385	093115	093115 ascaris lum
2	1575	77.7	390	09XYU9	09XYU9 wuchereria
3	1382.5	68.2	389	019707	019707 caenorhabdi
4	1310.5	64.7	433	018479	018479 meloidogyne
5	725.5	35.8	364	09XYM7	09XYM7 caenorhabdi
6	662	32.7	127	096775	096775 brugia mala
7	646	31.9	290	021808	021808 caenorhabdi
8	602.5	29.7	550	021540	021540 caenorhabdi
9	595	29.4	484	019053	019053 caenorhabdi
10	578.5	28.6	315	093532	093532 caenorhabdi
11	546.5	27.0	384	022164	022164 caenorhabdi
12	535	26.4	403	090250	090250 caenorhabdi
13	511	25.2	610	09XYN2	09XYN2 caenorhabdi
14	477.5	23.6	440	093967	093967 caenorhabdi
15	476	23.5	349	023097	023097 caenorhabdi
16	451	22.3	609	022680	022680 caenorhabdi

17	432.5	21.3	731	5	009586	009586 caenorhabdi
18	414	20.4	382	5	018213	018213 caenorhabdi
19	393	19.4	647	5	021573	021573 caenorhabdi
20	376.5	18.6	477	5	09T2E2	09T2E2 caenorhabdi
21	361	17.8	66	5	096776	096776 brugia paha
22	357.5	17.6	437	5	019304	019304 caenorhabdi
23	339	16.7	225	5	061816	061816 caenorhabdi
24	304.5	15.0	789	5	09BRQ4	09BRQ4 caenorhabdi
25	293	14.5	366	5	022627	022627 caenorhabdi
26	290.5	14.3	344	5	045890	045890 caenorhabdi
27	289.5	14.3	1286	5	09TXR6	09TXR6 caenorhabdi
28	244	12.0	809	5	09U3W7	09U3W7 caenorhabdi
29	202.5	10.0	711	5	09NHV4	09NHV4 caenorhabdi
30	202.5	10.0	711	5	022724	022724 caenorhabdi
31	199	9.8	462	5	09WI43	09WI43 drosophila
32	174	8.6	284	5	09G0H4	09G0H4 caenorhabdi
33	166	8.2	604	5	09VZC5	09VZC5 drosophila
34	163	8.0	1549	5	09BMD4	09BMD4 drosophila
35	163	8.0	1557	5	09BMD5	09BMD5 drosophila
36	163	8.0	1638	5	09VEX3	09VEX3 drosophila
37	162.5	8.0	611	5	09VZE0	09VZE0 drosophila
38	151.5	7.5	514	5	016304	016304 caenorhabdi
39	150	7.4	758	5	09V9X1	09V9X1 drosophila
40	148.5	7.3	362	5	0960V7	0960V7 drosophila
41	142	7.0	833	5	09V9X0	09V9X0 drosophila
42	141	7.0	3507	5	023587	023587 caenorhabdi
43	140	6.9	692	5	09W5Z0	09W5Z0 drosophila
44	137	6.8	1262	5	018117	018117 caenorhabdi
45	135.5	6.7	744	5	09VAG2	09VAG2 drosophila

ALIGNMENTS

RESULT 1
ID 093115 PRELIMINARY; PRT; 385 AA.

AC 093115;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE CUT-1-like cuticlin protein precursor.

GN ASCUT-1.
OS Ascaris lumbricoides (common roundworm).
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Ascarididae; Ascaridoidea;
OC Ascarididae; Ascaris.
OX NCBI_TaxID=6252;

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97390131; Pubmed=9249070;
RA Timlinouni M., Bazicalupo P.;
RT "cut-1-like genes of Ascaris lumbricoides";
RL Gene 193:81-87(1997).
DR EMBL: U73005; AAB6646.1; -;
DR InterPro: IPR001507; Endoglin/CD105.
DR SMART; SM00241; ZP; 1.

KW Signal.
FT SIGNAL 17
FT CHAIN 17 385
SQ SEQUENCE 385 AA; 42408 MM; 103C6834BID29286 CRC64;

Query Match 81.6%; Score 1653.5; DB 5; Length 385;
Best local Similarity 80.2%; Pred. No. 1.5e-151;
Matches 311; Conservative 32; Mismatches 40; Indels 5; Gaps 4;

QY 2 MIMRIACFTLIALSYIPDNGVEGEPEIECGFTSTINPNRNATGEHYVYKGLYDQ 61
DB 1 MCRAYSF-LALFGIAAIPVDNGVEGEPEIECGFTSTINPNRNATGEHYVYKGLYDQ 59
QY 62 GCRNDEGROVAGISLEPDSNCVARTSLNPRGIFVTTVVISFHPLEFVKVBRAYVOC 121
DB 60 GCASDEGROVAGISLEPDSNCVARTSLNPRGIFVTTVVISFHPLEFVKVBRAYVOC 119

```

QY 122 FYNEADKTVSAQIEVSEITTAFTQTOIVMPVCRREIILDGPGTGPQVFAITGQPVYHKWT 181
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 120 FYNEADKTVSAQIEVSEITTAFTQTOIVMPVCRREIILDGPGTGPQVFAITGQPVYHKWT 179
QY 182 CDESEVDFCAVYHSCFVDGNGDTEVIELNADGALDKYLLNNEVPTDLMAGQEAHVYK 241
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 180 CDESEVDFCAVYHSCFVDGSGDTQIILNEGCALDKYLLNNEVPTDLMAGQEAHVYK 239
QY 242 YADRSQLFYQCOISITIKENPSECVPOGSEPOGFCAGVKTGA-AAKPAALALRLKKR 300
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 240 YADRSQLFYQCOISITIKENPSECVPOGSEPOGFCAGVKTGA-AAKPAALALRLKKR 299
QY 301 SAEPENIIDVRDINTLEISDNOALPVDLRHALLQ-NGQPVILAAVONGICMSPFGF 359
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 300 GGQYDNTLDVPTDFSAIDISDRDEALPMDLRHRA--RHARGQOYILSPANEGICMSPFGF 357
QY 360 SMFGLSIALIAVAIITISFKFRPNOKA 387
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 358 SIMGLAVLAAVAVVVSFKLRPOKA 385

```

```

RESULT 2
Q9XYU9 PRELIMINARY; PRT; 390 AA.
AC Q9XYU9;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Cuticulin-1.
GN CUT-1.
OS Wuchereria bancrofti.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
OC Onchocercidae; Wuchereria.
OX NCBI_Taxid=6293;
RN [1]
RP SEQUENCE FROM N.A.
RA Ramzy R., Helmy H., Adely M., Curtis K., Well G.;
RT "Wuchereria bancrofti L3 cuticulin-1 cDNA.";
RL Submitted (JUL-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF125580; M28743.2;
DR InterPro; IPR001507; Endoglin/CD105.
DR SMART; SM00241; 2P; 1.
SQ SEQUENCE 390 AA; 43026 MW; 2CA1E7E3D39FRA5E CRC64;

```

```

Query Match 77.7%; Score 1575; DB 5; Length 390;
Best Local Similarity 78.8%; Pred. No. 6.2e-144;
Matches 301; Conservative 29; Mismatches 44; Indels 8; Gaps 4;

```

```

QY 2 MRLIAFCTTLALSYIPVNDGVEGEPEIEGPGTSTITNFTRNAFEHVVYKGLYDOE 61
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1 MRLISILVAFTTIVAIIPVNDGVEGEPEIEGPGTSTITNFTRNAFEHVVYKGLYDOE 60
QY 62 GCNDEGGRVAGISLPFSCNVAFTRSINPRGIEFTTVVISFHLFTKVDRAVYOC 121
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 61 GCNDEGGRVAGISLPFSCNVAFTRSINPRGIEFTTVVISFHLFTKVDRAVYOC 120
QY 122 FYNEADKTVSAQIEVSEITTAFTQTOIVMPVCRREIILDGPGTGPQVFAITGQPVYHKWT 181
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 121 FYNEADKTVSAQIEVSEITTAFTQTOIVMPVCRREIILDGPGTGPQVFAITGQPVYHKWT 180
QY 182 CDESEVDFCAVYHSCFVDGNGDTEVIELNADGALDKYLLNNEVPTDLMAGQEAHVYK 241
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 181 CDESEVDFCAVYHSCFVDGNGDKVELLNADGALDKYLLNNEVPTDLMAGQEAHVYK 240
QY 242 YADRSQLFYQCOISITIKENPSECVPOGSEPOGFCAGVKTGA-AAKPAALALRLKKR 300
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 241 YADRSQLFYQCOISITIKENPSECVPOGSEPOGFCAGVKTGA-AAKPAALALRLKKR 300
QY 301 SAEPENIIDVRDINTLEISDNOALPVDLRHALLQ-NGQPVILAAVONGICMSPFGF 358
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 301 DVRDENIVDVRDINTLEISDNOALPVDLRHALLQ-NGQPVILAAVONGICMSPFGF 358
QY 359 FSM-----FMGLSIALIAVAIIT 375

```

```

Db 361 FTLAGMLIEFVYSAVATVATL 382
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

```

RESULT 3
Q19707 PRELIMINARY; PRT; 389 AA.
AC Q19707;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE F22B5.3 protein.
GN F22B5.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Sims M.A.;
RL Submitted (JUL-1995) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C. elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; Z50044; CAA90355.1;
DR InterPro; IPR001507; Endoglin/CD105.
DR SMART; SM00241; 2P; 1.
SQ SEQUENCE 389 AA; 42983 MW; 40BAB9C5CE031BB5 CRC64;

```

```

Query Match 68.2%; Score 1382.5; DB 5; Length 389;
Best Local Similarity 69.6%; Pred. No. 2.7e-125;
Matches 268; Conservative 42; Mismatches 52; Indels 23; Gaps 8;

```

```

QY 12 LIALSYIPVNDGVEGEPEIEGPGTSTITNFTRNAFEHVVYKGLYDOEGRNDEGGR 71
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 13 LVASVAIIPVNDGVEGEPEIEGPGTSTITNFTRNAFEHVVYKGLYDOEGRNDEGGR 72
QY 72 VAGISLPFSCNVAFTRSINPRGIEFTTVVISFHLFTKVDRAVYOCFMEADKTVS 131
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 73 VAGISLPFSCNVAFTRSINPRGIEFTTVVISFHLFTKVDRAVYOCFMEADKTVS 132
QY 132 AQIEVSEITTAFTQTOIVMPVCRREIILDGPGTGPQVFAITGQPVYHKWTCDSEVDFPC 191
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 133 TQIEVSDLTATQTOVYPPRICKYELNNGPTGEPYQFATIGQYVYHKWTCDSEVDFPC 192
QY 192 AVVHSCFVDGNGDTEVIELNADGALDKYLLNNEVPTDLMAGQEAHVYKADRSQLFYQ 251
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 193 AVVHSCFVDGNGDTEVIELNADGALDKYLLNNEVPTDLMAGQEAHVYKADRSQLFYQ 252
QY 252 CQISITIKENPSECVPOGSEPOGFCAGVKTGA-AAKPAALALRLKKRSAE-PENIIDY 310
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 253 CQISITIKENPSECVPOGSEPOGFCAGVKTGA-AAKPAALALRLKKRSAE-PENIIDY 307
QY 311 RFDINTLEISDNOALPVDLRHALLQ-NGQPVILAAVONGICMSPFGS--FMFG 364
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 308 RABLTLEIELEGN--LPSSLTQOALVASRELGED----SFQDELICISSFHLISVTVLELG 361
QY 365 LSTALIAVAIIT--ISFKFRPNOK 386
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 362 LTV--FVALFTYMTIVSRMVPSPDK 384

```

```

RESULT 4
O18479 PRELIMINARY; PRT; 433 AA.
ID O18479;
AC O18479;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)

```


QY 356 P-----EFSMFPGISTALIAVITITISFKRPNO 385
 DB 342 PKLLVAVFETFEFLFVLTITLVVYHRYCKNE 375

RESULT 12

Q90250 ID Q90250 PRELIMINARY; PRT; 403 AA.

AC 090250; 01-MAY-2000 (TReMBLrel. 13, Created)
 DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
 DE 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
 GN Y53H1B.1 protein.

OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxId=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA White S.;
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode C.elegans: A platform for
 RT investigating biology.";
 RL Science 282:2012-2018(1998).
 DR EMBL; AL132851; CAB60411.1; -;
 DR InterPro; IPR001507; EndoglIn/CD105.
 DR SMART; SM00241; ZP; 1.
 SQ SEQUENCE 403 AA; 45346 MW; EECSEB5419CA8428 CRC64;

Query Match 26.4%; Score 535; DB 5; Length 403;
 Best Local Similarity 31.0%; Pred. No. 3.2e-43;

Matches 126; Conservative 77; Mismatches 160; Indels 44; Gaps 13;

QY 9 CTTLLALSYSI-----PVNDGVEGEPEIECGPTSTINFTNRAFGHYVKGLYDDEG 62
 DB 5 CYLLVLLITGVMAADPPTLDNGISELPEVDCMDRYKLSKTYORPHGRIFVGVGMVDKA 64
 QY 63 CRNDEGROVAGISLPFD--SCNVAFTSLNP--RGIFVTTVVISFHPLEFVTKVDRAV 118
 DB 65 CVADFTYSAQKDYTFELENGACNMRQRMIGPEKGMESMVIYISFHSFTITKVDRAV 124
 QY 119 VQCFWEAKTVSAQIEVSEITTAFTQTVPMPCRYEILDGPTGQPVQAFALIGPVYH 178
 DB 125 CTCEFYEMADKVVYNNKFPDMLPTDLDITARMPLCTYSVRDSITGPIVEFAKVGELVYH 184
 QY 179 KMTCDSEFYDTFCVAVHSCFVDGNGDTVE--ILNADGCAIDKYLNNLEYPFD--LMAQOE 236
 DB 185 VNNCES--DMSMLVHSCFVDGNGDERKPLIDEGHCALDPLILDITLYNKDNVAYAQ 241
 QY 237 AHVYKADRSQLEFYCOQISITTIKEPNSC---VRQCSSEPGGAVTKGGAAPRAAAQ 293
 DB 242 VNTFKFADRVSYTFQCAVS--TCMNTGCMCGKTPRCGPAGSP-----RSSSSSN 290
 QY 294 LRLTKRSAPEPNIIDVRDINTLEISDD-----NQALVVDLRLHALLHNGQPVY---L 345
 DB 291 DNGFNRRSRA--TLNKRVAVMVHLADTMDLANKITVPELEEKNSDCKDEPRAHSHL 348
 QY 346 AAVONG-----ICMSPFGSMFGISIALIAVITITISFKRPNO 387
 DB 349 ALIQGSHPTVYVCLNGSRIOIFVLSFLMLTVTVTLITTVRSKMA 395

RESULT 13

Q9XVNZ ID Q9XVNZ PRELIMINARY; PRT; 610 AA.

AC 09XVNZ; 01-NOV-1999 (TReMBLrel. 12, Created)
 DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)

DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
 DE F53B6.6 protein.
 GN F53B6.6.

OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxId=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA White S.;
 RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode C.elegans: A platform for
 RT investigating biology.";
 RL Science 282:2012-2018(1998).
 DR EMBL; Z81086; CAB03116.1; -;
 DR InterPro; IPR001507; EndoglIn/CD105.
 DR SMART; SM00241; ZP; 1.
 SQ SEQUENCE 610 AA; 68585 MW; 09819DC5C3741329 CRC64;

Query Match 25.2%; Score 511; DB 5; Length 610;
 Best Local Similarity 37.4%; Pred. No. 1.2e-40;

Matches 101; Conservative 58; Mismatches 101; Indels 10; Gaps 6;

QY 10 TLLIALSYIS--IPVNDGVEGEPEIECGPTSTINFTNRAFGHYVKGLYDDEG--RN 65
 DB 61 TLLIILSTSSCFEIONGVGKPEVFCGIDTINKVNTFENGRITVDESKKOCVOHS 120
 QY 66 DEGGRVAGISLFPDSCNVAFTSLNP-RGIFVTTVVISFHPLEFVTKVDRAVVOCFYWE 125
 DB 121 ADAHSSPOEFTIPDGACNNRQRTLHPRGISFTMTISFHFYTGMDRAISINCFLE 180
 QY 126 ADKTVSAQIEVSEITTAFTQTVPMPCRYEILDGPTGQPVQAFALIGPVYHKTCDSE 185
 DB 181 SIKGLNAEIDVGTLPAPQHVDOEYSLPVCAYHMKD--GIEGHVLFRAOVGGKYVHWRCDD 239
 QY 186 TYDTFCVAVHSCFVDGNGDVTIELINADGCAIDKYLNNLEYPFD--LMAQOAHVYKXAD 244
 DB 240 ASHVYGLILHSCYADGDHKNKRELVDGCSYDPLLPQIEYEHGAISAYTNAHVRYAD 299
 QY 245 RSQLEFYCOQISITTIKEPNSC---VRQCS 271
 DB 300 KVQLYFTCTVQLCYKH--DGCCEGITRPQCS 328

RESULT 14

Q93967 ID Q93967 PRELIMINARY; PRT; 440 AA.

AC 093967; 094405; 01-FEB-1997 (TReMBLrel. 02, Created)
 DT 01-NOV-1999 (TReMBLrel. 08, Last sequence update)
 DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
 DE ZK265.8 protein.

GN ZK265.8.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxId=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Dobson R.;
 RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier W., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,

RL Science 282:2012-2018(1998).

Db 34 PKLEGGSEGILKIHINPTGFGHHYVVRGFEPPQTVCHLNYCTRLNRPIMVDLPFRGPCNV 93

Qy 85 ARTBSLNPRGIEVTITVVISFHPLEVTKVDRAYRVQCYMEADKTVSAQIEVSEI-TTAF 14

Db	94	RRRRNVAPSSISYDVTVIILIIHNPRLFTSPDRAKRLNLCIRQDESLTQQRINNSDIPSTAL	153		
Qy	144	QTOIVMPRCRYEILIDGGPTGQPVQFALIGQRPYHAKMTCDSFTVDFCAVHSCFVDDGN	203		
Db	154	QSKNAPK--CRYDVLGSSINGVVRFRANQGDVYVHKMTQDS--DRFGFVHSCVVRDSS	208		
Qy	204	GDTVELNADGCAALDKYLLINLEFPFDLMAGDEA-HVYYADRSQLFYCCQISITIKERN	262		
Db	209	GKDFPFIIDRGCVTFDFSLFEVYSDDLSAFTAVAFRAFYADQVMVHFSCQILTTCQKORN	268		
Qy	263	S-ECVYRPOCSEPFQFGAVK	280		
Db	269	GCEGISPPICRPMDUGPIK	287		
RESULT 17					
ID	Q09586	PRELIMINARY:	PRT: 731 AA.		
AC	Q09586:				
DT	01-NOV-1996 (TREMBLrel. 01, Created)				
DT	01-OCT-2001 (TREMBLrel. 18, Last sequence update)				
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)				
DE	Hypothetical 81.9 kDa protein.				
GN	K06A1.3.				
OS	Caenorhabditis elegans.				
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;				
OC	Rhabditidae; Pelodermidae; Caenorhabditis.				
OX	NCBI_TaxID=6239;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-BRISTOL N2;				
RX	MEDLINE=99069613; PubMed=9851916;				
RA	None;				
RT	"Genome sequence of the nematode C. elegans: a platform for				
RT	investigating biology. The C. elegans Sequencing Consortium."				
RL	Science 282:2012-2018(1998).				
	[2]				
RL	"The sequence of C. elegans cosmid K06A1."				
RL	Submitted (APR-1995) to the EMBL/Genbank/DBJ databases.				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-BRISTOL N2;				
RA	Waterston R.;				
RT	"Direct Submission."				
RL	Submitted (JUN-2001) to the EMBL/Genbank/DBJ databases.				
DR	EMBL; U23449; AAC24298.2; -.				
SW	Hypothetical protein.				
QY	SEQUENCE 731 AA; 81860 MW; C7D6FB6864054945 CRC64;				
Query Match 21.3%; Score 432.5; DB 5; Length 731;					
Best Local Similarity 33.4%; Pred. No. 5.9e-33;					
Matches 96; Conservative 53; Mismatches 111; Indels 27; Gaps					
Qy	6	IAFTCTLLALSYS---IPVNGVGEPELEICGPTSIINFTRNAFEGHYVKGGLYDOE	61		
Db	4	IDFLILMALRLNSAQPIYEFNDHISDEPEVSHSGFMGLKVTESPSHYVKGKFRKD	63		
Qy	62	GCRNDEGGROYAGISLPDSCNVAFTSLNRPGRIVTTTVVISFRLPVTYKDAYRYOC	121		
Db	64	GC---SFSNANATFEESKCDVMQROBANPGRMAYTATVAVQVQLPFTTKDAYKRLKC	119		
Qy	122	FYMEADKIVNSQIIESETITTAFOQIIVMPRCRYEILIDGGPTGQPVQFALIGQRPYHAKT	181		
Db	120	FYKEAEKAVGAEVSDPTPVQLEDESPQVPCSYTIHKESSPGRIAKFAQLGDVLYHWE	179		
Qy	182	CDSEITDFCAVHSCFVDDGNQDVELILNMDGALDKYLLINLEFPFDLMAGDEAHY	240		
Db	180	CPSE---AYQMEVINCVDVYGGEEYSKKVIIGENGCSEDITIMENLITNERTKAFVNSNAF	236		
Qy	241	KYADRSOLFQCCQISI--TIKEPNSECVRPC-----SEPG	275		

```

Db      237 NFPDQNNVRISCKIYVCPTL---SGTCHOPKPCDASASIDDASEPIG 280
          : : : : : | : | : : : | : : : : | : | : |
RESULT# 18
ID       018213 PRELIMINARY; PRT; 382 AA.
AC       018213;
DT       01-JAN-1998 (TREMBLrel. 05, Created)
DD       01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DE       01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE       Y53C12B.6 protein.
GN       Y53C12B.6.
OS       Caenorhabditis elegans.
OC       Eukaryota; Metazoa; Nemtoda; Chromadorea; Rhabdilitda; Rhabditoidea;
OC       Rhabdilitdae; Peloderinae; Caenorhabditis.
OX       NCBI_Taxid=6239;
RN       [1]
RP       SEQUENCE FROM N.A.

Query Match           20.4%; Score 414; DB 5; Length 382;
Best Local Similarity 31.6%; Pred. No. 1,6e-31;
Matches 86; Conservative 56; Mismatches 112; Indels 18; Gaps

Oy      21 VDNVGEPEEICGPTSTINFTNFAEGHYVVKGLYDOECRNDGGRQVAGISLP-- 78
         |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      17 VSDIIIGEKIKCAPGITITMETDSPFGALFLRSADKSKCANFSAQPSQNTSIFERG 76
         ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

Oy      79 FDCSNVARTSL-NPGLIEVTITTVVISFHPLFVTKVDRAARYOCFYMEADKTVAQIEVS 137
         ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      77 FDDCSRRKRQIVABRGWMSVLVYSYHGSILTHRDVAYQIDCFYRENSRVETMLSVN 136
         ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

Oy      138 ETTTFAPQGIYMPRCRYEI-LDGER-----TGPRQFALIGPYVHNKPTCDS 185
         |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      137 ADOPRLISEPLPLPCDRVEVTGAAGAAGIVTSLSLTASOIANYGSGVIHWTCSD 196
         |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

Oy      186 T-VDFPCAIVHSCFPDDNGDGFVEILLMDGCALDKYLLNLLEYEPIDLM-AGEAHYYKYA 243
         |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      197 AASDVYCIGVYSTCAEDGDSDIVQYVDENCGCTTDGBELLSPYKKGSMMAAASHAFKV 256
         |:|:~::~||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

Oy      244 DRSQLFYQCQSITTIKEPNSECVRPCQCEPG 275
         |:|:~::~||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      257 DNHIYVFKNIRITVKNPSGECPPVNCSS-PNG 287
         |:|:~::~||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

RESULT# 19
ID       021573 PRELIMINARY; PRT; 647 AA.
AC       021573;
DT       01-NOV-1996 (TREMBLrel. 01, Created)
DD       01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE       01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE       M28.1 protein.
GN       M28.1.
OS       Caenorhabditis elegans.
OC       Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitda; Rhabditoidea;
OC       Rhabdilitdae; Peloderinae; Caenorhabditis.
OX       NCBI_Taxid=6239;
RN       [1]
RP       SEQUENCE FROM N.A.
```



```

RESULT 22
ID Q19304 PRELIMINARY; PRT; 437 AA.
AC Q19304;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DE 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Hypothetical 49.1 kDa protein.
GN F10E7.10.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditolea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Pauley A.;
RT "The sequence of C. elegans cosmid F10E7.";
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; U41264; AAA82426.1;
DR InterPro; IPR001507; Endoglin/CD105.
DR SMART; SM00241; ZP; 1.
KW Hypothetical protein.
SQ SEQUENCE 437 AA; 49113 MW; 033C2B12BFCB748D CRC64;

Query Match 17.6%; Score 357.5; DB 5; Length 437;
Best Local Similarity 31.6%; Pred. No. 5.5e-26;
Matches 87; Conservative 47; Mismatches 122; Indels 19; Gaps 7;

QY 1 MMRLAFTCTLLALSIYVDNGVESEPEIEGGPTISITINENTRNAFEHVVYKGLYDQ 60
DB 1 MLISIVWYSFFITVWSSLYKPEIVDTPIVCEEDKTIITKTSMSNPSTNITVDEGREN 60
QY 61 EGGRNDEGGROVAGISLPEDSCNVAARTSLNPRGIFVTTTVISFHLPTKVDRAVRVQ 120
DB 61 MECVS---RNQKIEVAHDKCGVANEKTEQRPNGSTRLCIFQVLPHPVTESDRSYCAQ 116
QY 121 CFYMEADKIVSAQIEVS-EITTAFOFOIYP-----MPVCRYEILDGGPTGQPVQFAITG 173
DB 117 CVY-ADSHVMKDIESTLLISEAPQLQSPQDAPVMPKCNYSIRKQKGPVQYASIG 174
QY 174 QPYVHKMTDSEIVDFCAVNHSCFYDDGNGDTVELINADGALDKYLLNLEYPTD-LM 232
DB 175 DSVFHHWSCDGNHNG---ILVQNCHEVDQGNKILLIIDNGCGIDHYVMDPTLYNGEOSI 231
QY 233 AGQEAHVYKYADRSOLFYOQISITIKPEPNSGCVR 267
DB 232 AFQETHVFKEFAQRTVTRFICQIKTKCK--GDCCR 264

RESULT 23
ID 061816 PRELIMINARY; PRT; 225 AA.
AC 061816;
DT 01-AUG-1998 (TReMBLrel. 07, Created)
DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)
DE 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Hypothetical 25.2 kDa protein.
GN B0511.5.

```

```

OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditolea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Tin-Mollam A., Sutterer C., Ozersky P.;
RT "The sequence of C. elegans cosmid B0511.";
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF067608; AAC17653.1;
DR InterPro; IPR001507; Endoglin/CD105.
DR SMART; SM00241; ZP; 1.
KW Hypothetical protein.
SQ SEQUENCE 225 AA; 25235 MW; 29070A6340C1292C CRC64;

Query Match 16.7%; Score 339; DB 5; Length 225;
Best Local Similarity 31.2%; Pred. No. 1.4e-24;
Matches 70; Conservative 42; Mismatches 108; Indels 4; Gaps 4;

QY 49 EGHVYKGLYDQEGCRNDEGGROVAGISLPEDSCNVAARTSLNPRGIFVTTTVISFHL 108
DB 4 KGRVFLVGHSDKDCVSRERGRRTTITVPRDKGVETVQHGKAGAYTSVNTVISFHDK 63
QY 109 FVTVKDVRAVRVQCFYMEADKIVSAQIEVSEITTAFOFOIYVMPVCRYEILDGGPTGQPVQ 168
DB 64 FLTRVRAVRVITGLVAPTGQVVSXALTVQPSLLKDIQVLAEGSCSEYFV-D-VTRRPAE 122
QY 169 FAIIQGPVYHKMTCDSEIVDFCAVNHSCFYDDGND-IVELINADGALDKYLLNLEY 227
DB 123 IYHVNAPLEHVTCDSTNDLDFCMYHDCVINEGSKRSKLIIDSGCSLDITTRLPLMY 182
QY 228 PTD-LMAQGEAHVYKYADRSOLFYOQISITIKPEPNSGCVRPOC 270
DB 183 ENNKLSARVMSKAFREGDVAVVEFCNVRLDLRNGTS-CPRPRC 225

RESULT 24
ID 09BK04 PRELIMINARY; PRT; 789 AA.
AC 09BK04;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DE 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Hypothetical 89.1 kDa protein.
GN Y37811A.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditolea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]

```

RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Ozersky P., Stromatt C.;
RT "The sequence of C. elegans cosmid Y37B11A."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RT "Direct Submission";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC091267; AAK31565.1; -
DR InterPro: IPR001507; Endoglin/CD105.
DR InterPro: IPR002035; VWF_A.
DR Pfam: PF00092; vwa; 2.
DR PRINTS: PRO0453; VWFADOMAIN.
DR SMART; SM00327; VWA; 2.
DR SMART; SM00241; ZP; 1.
DR PROSITE; PS50234; VWF_A; 2.
KW Hypothetical protein.
SQ SEQUENCE 789 AA; 89078 MW; 1E930119092A5334 CRC64;

Query Match 15.0%; Score 304.5; DB 5; Length 789;
Best Local Similarity 26.4%; Pred. No. 1.6e-20;
Matches 78; Conservative 58; Mismatches 117; Indels 43; Gaps 12;

OY 25 VEDEPELECGPTSTITNFNTRNAPFEGHYVYKGLYDQEGC--RDDEGRQVAGISLPDSC 82
DB 445 VRGATLVKCTEHSMTIVRQRLQGVMAHMYHDEPECIRRTDINSREIQ-MTFEGKC 503
OY 83 NVARTSLNRPGLFVTTTVIVISFPLFVTKVDRAYVQCQYMEADKTVASQIEVSETTTA 142
DB 504 GLVKTPTADGCHGNFNTVILQFHPRLITRADGLDKSCVSSA--VPRDELRAVYLKNA 561
OY 143 FQTOIVPMVRCRYEILDGPTGPVOF-ALIGQPVYHKWTCDETVDFCAVYHSCFVD 201
DB 562 ADTQ-----CYVRLHRYSP-GQCVALDAKVGETLYHRMADSDSPREYNY--LVHDFCV-Q 611
OY 202 GNDQYVILLNADGCALDKYLLNN-----LEYPTDLMAQGEAHYKYKADSQLFYQOQIS 255
DB 612 SEKHQOILLDSNGCEVDQHELETPNYSRFKDYPEDSYVFEKMSVFRPPGDGDLFFHCKIS 671
OY 256 I-TIKENPSEC--VRPOCSEPOGFGAVKTGGAAPAAARLRLKRSAPENI 307
DB 672 LCMANDPNAPCNSIIPKC-----PKKPVLPVROKRSVSAPEM 710

RESULT 25
O22627
ID 022627 PRELIMINARY; PRT; 366 AA.
AC 022627;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE T21B10.6 protein.
GN T21B10.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBL_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Baynes C.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C. elegans: A platform for
investigating biology."
RL Science 282:2012-2018(1998).
DR EMBL; Z68318; CAA82696.1; -
DR InterPro: IPR001507; Endoglin/CD105.

DR SMART; SM00241; ZP; 1.
SQ SEQUENCE 366 AA; 40916 MW; 38D248E42F0187F5 CRC64;

Query Match 14.5%; Score 293; DB 5; Length 366;
Best Local Similarity 24.6%; Pred. No. 7.7e-20;
Matches 81; Conservative 53; Mismatches 121; Indels 74; Gaps 10;

OY 75 ISLPDSCNVARPRLNPRGLFVTTTVIVISFHPFVTKVDRAYVQCQYMEADKTVASQI 134
DB 63 ILIPHECMVPRRSLHPGTGVLVSLVSFHPFETTVDDRIEMQCFHKKINGTSTSL 122
OY 135 EVSEITTAFOQIYPM-----PVCRYEILDGPTGPVOFAIIGQPVYHKWTCDESET 186
DB 123 AIGS-----PKRPPSDTKGPSCSYEVLV-SPEGLRAGRALGQDYYHSMQC--QN 169
OY 187 VDFCAVYHSCFVDDGNGDTVEIILNADGCALDKYLLNNLEYPTDLMAQGEAHYKYKADRS 246
DB 170 VYESCTMIENCELVGGE-ETHNEVIDSSGCKNHEIMQLEYNHRTVGTGVKGVSHTS 228
OY 247 QLFYQOQISTITKEPNSECVRPOCSEPOGFGAVKTGGAAPAAARLRLKRSAPEN 306
DB 229 IVYFACQVRLHPOLPTGECRPKC-----DLMRKRKEDSSQFP 266
OY 307 IIVRPTDINTLEISDDNQALPVDLRHALLQHNQGPVI-----LA 346
DB 267 SIDVRS--QNLEISQ-----LINTSSTMPHPTEPPIQHTCPDINHETESIERASKA 319
OY 347 AVONGICMSPFQSMFNGSLIALAVII 375
DB 320 LDEBRICAD--FRSVLWVSILLTVALIL 345

RESULT 26
O45890
ID 045890 PRELIMINARY; PRT; 344 AA.
AC 045890;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE W06D12.1 protein.
GN W06D12.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBL_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Basham V.M.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C. elegans: A platform for
investigating biology."
RL Science 282:2012-2018(1998).
DR EMBL; Z682073; CAB04922.1; -
DR InterPro: IPR001507; Endoglin/CD105.
DR PRINTS; PR01651; SECSEXPORT.
DR SMART; SM00241; ZP; 1.
SQ SEQUENCE 344 AA; 38563 MW; D3D50114B7B4891A CRC64;

Query Match 14.3%; Score 290.5; DB 5; Length 344;
Best Local Similarity 27.4%; Pred. No. 1.2e-19;
Matches 82; Conservative 45; Mismatches 87; Indels 85; Gaps 12;

OY 151 PVCRYEILDGPTGPVOFAIIGQPVYHKWTCSETVDFCAVYHSCFVDDGNGDTVEIL 210
DB 20 PVCRYEILMNAQGPRLSHATVGDLYHKMSCGNKEMKCMYHSCVVDGQGFQGLV 79
OY 211 MADGCALDKYLLNNLEY-PTDLMAQGEAHYKYKADRSQLFYQOQISTITKE-----P 261
DB 80 DEHGCTLDATILKELEYNEKELEAGQMSVFKFADKPTVFPSCMIVREKESAMPCVIP 139

```

QY 262 NSECVRPPCCSE-----PQGGVAVKGTGAANKPAAAOQLRLKKRSA---- 301
      | : |
Db 140 TEVCKNIPSSKQMLNDVTDKRVSDDIPGF-----PRPNNTKLOKSQKSFSDAM 190
      | : |
QY 303 -----EPENITID-----VRTDI--NTLEISDQNALP----VD 321
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 191 TSNDSYADFLDDELEDEVDKMSRIPTYTNMRLIRVDVAGNTKILAVDVAAAPSVNYQD 250
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 330 LRRHALLQHGNGOPV-----ILAAVQNICGSPGFSMEGMSIALIILAA--IITISFEK 381
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 251 LPPESEV---GNPQNGKREHFVPTNAASNDVCTFRIN---LILSLIILAAIGDIVITISLKY 301

```

```

RESULT 27
ID Q9YXR6 PRELIMINARY; PRT; 1286 AA.
AC Q9YXR6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE M01E10.2 protein.
GN M01E10.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT Investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX Gettel C., Kramer J., Smith A.;
RT "The sequence of C. elegans cosmid M01E10.";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX Waterston R.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF099921; AAC68807.1;
DR InterPro: IPR001507; Endoglin/CD105.
DR InterPro: IPR002965; P_rich_extensn.
DR InterPro: IPR002035; VWF_A.
DR Pfam; PF00092; vwa; 1.
DR PRINTS; PRO1217; PRICHEXTENS.
DR PRINTS; PRO0453; VWFADOMAIN.
DR SMART; SM00327; VWA; 1.
DR SMART; SM00241; ZP; 1.
DR PROSITE; PS50234; VWF_A; 1.
SQ
SEQUENCE 1286 AA; 140794 MW; 95781999BC127CCE CRC64;
Query Match 14.3%; Score 289.5; DB 5; Length 1286;
Best Local Similarity 28.0%; Pred No. 8.7e-19;
Matches 97; Conservative 46; Mismatches 156; Indels 46; Gaps
QY 30 EIECGPTSLINFNTRNAEGVHYVKGLYDDEGCRNDEGGRVAGISLP-----DSCN 83
Db 907 DLECVGGEGFVQVNNPAGFGVAVVAGCYDDARC-----KATASSSLPILNFISSNCG 960
QY 84 VARTSLNPGIVTITVVIVISEHFLFVTKVDRAIRVQCFY--MEADKIVSAQIEV--SEI 139
Db 961 VTQVKKSSDPKGLNSIVLHLHDDLENTAEDRAVYLQCFIGAQNODAVSTNLNVRSLE 1020
QY 140 TTAFOQIIVM-VVCVKEILLDGPTGQVQFAIIGQVYHNKTCQ--SEYDTCAVYVS 196
Db 1021 AIAETISLSTLPTCTYSLRKREPERIVSRVAVGQVWNRNESCGLTDQAGVIOVNS 1080

```

[illegible]

RESULT	ID	PRELIMINARY;	PRT;	809 AA.
0903W7	Q9U3W7			
AC	Q9U3W7;			
DT	01-MAY-2000 (TREMBLrel. 13, Created)			
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)			
DT	01-MAR-2002 (TREMBLrel. 20, Last annotation update)			
DE	SP460.			
GN	SP460.			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC	Ephydroidea; Drosophilidae; Drosophila.			
OX	NCBI_taxid=7227;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Sereno T.L., Pendleton J.D., Rubin G.M.;			
RT	"A reverse genetic screen for genes involved in Drosophila			
RT	development."			
RL	Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AF214523; AAF24502.1; -			
DR	FlyBase; FBgn0029129; SP460.			
DR	InterPro; IPR000561; EGF-like			
DR	InterPro; IPR001507; Endoglin/CD105.			
DR	SMART; SM00181; EGF; 5.			
DR	SMART; SM00241; ZP; 1.			
DR	PROSITE; PS01186; EGF_2; UNKNOWN_4.			
DR	SEQUENCE 809 AA; 87817 MW; 07FD2F0445FD487C CnC64;			
Query Match		12.0%;	Score 244;	DB 5; Length 809;
Best Local Similarity		23.7%;	Pred. NO. 1.2e-14;	
Matches 100; Conservative		54;	Mismatches 162;	Indels 106; Gaps 20;
QY	7 AECTLLALSTIPVD--NGVEGPEETICGSPSTIINN-----TR 45			
Db	393 AFCN---AVNHRACQCCTITGTGNDLCHNHTNERTDEPRPDMVYVSLADGVQVEIHTE 449			
QY	46 NAFEGHVYKGLYDOEGGRNDEGGHQAIVGISL-----FDSGNVARTSLNRGITFV 97			
Db	450 PGFNQVLYVKHSHKSNDEGRRV---NLAGEIVPRFETIRVHFGSGMAVADV-----A 500			
QY	98 TTVTVVISFHPLEVTKVDRAVYVQCFYMEADKTVSAQIEVSEITTAQ--TQIVPMFCRYE 156			
Db	501 SEFVLVIQHPRLKYTKAQAQNYIKCYVQGEKNVTLGFNVSMLTGATIGTANGPPIQMR 560			
QY	157 ILDGGPIGQAPVQFALIGQPVYHKTKCSELYDTFCAYVHSCF---VDGNGDYIELNAD 213			
Db	561 IITN--EEELNSAIEIGNL--KLQVDVEPATIYIGGFARSCIARTKMEEDVONEYLYTDEN 616			
QY	214 GCAIDKYLNLNLEKPTD---LMAQGEHAVYVYADRSOLFQCCQISITIKEPNSECVRPQC 270			
Db	617 GCADTSTIFEGWMEYNPDNLSLASFN--EFKPPSDNIRFCQNTKVCF---GRQYPQNC 670			
QY	271 SEPQGFQAVTKGGAAPPAALAAQLRLKRSAREPENIIDVPTDINTLTLEISDNOALPVDL 330			
Db	671 GGYNAFGR-----RRRS-----IADNSTDMATA---TNSVVEGQL 703			
QY	331 RHRLALQHN-----GQPVV-----ILAAVQ--NGICMSPFGESEMGSLIALIAVIT 375			

Db 704 REBITSSNAITFEKRSKOGILANDANIKPAAQREVEDICVS-----MVGILIALVITALL 757
 QY 376 TI 377
 Db 758 TL 759

RESULT 29
 Q9NHV4 PRELIMINARY; PRT; 711 AA.

AC Q9NHV4; 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, last annotation update)
 DE Transmembrane protein RAM-5.
 GN RAM-5.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20359254; PubMed=10899109;
 RA Yu R.Y., Nguyen C.O., Hall D.H., Chow K.L.;
 RT "Expression of ram-5 in the structural cell is required for sensory
 ray morphogenesis in Caenorhabditis elegans male tail."
 RL EMBL J.19:3542-3555(2000).
 DR EMBL; AF218866; AAF67103.1;
 DR InterPro; IPR001507; Endoglin/CD105.
 DR SMART; SM00241; ZP; 1.
 DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; UNKNOWN_1.
 SQ SEQUENCE 711 AA; 80883 MW; 25F9FA06FE27E6 CRC64;

Query Match 10.0%; Score 202.5; DB 5; Length 711;
 Best Local Similarity 21.6%; Pred. No. 1.1e-10;
 Matches 63; Conservative 47; Mismatches 122; Indels 59; Gaps 10;

QY 5 LIAPCTLLIALSYIPVDNG-VESEPEIEGPGTSITINFTRNAFEHVVYKGLYDEGC 63
 Db 8 VIGIIFSVVLATSTDPDNAMHMGVPQVTCSAKLITVSFNTNIPFGCRISVFDKLTIPAC 67
 QY 64 RNDEGGROVAGISLPEDSCVARTSLNPRGIEFTTVVISFHLFTKYDRAVRVOCFY 123
 Db 68 NHDYSTNIQKNATFQMDILKCANPMLKNSRLRAYVEIGFHLVMTNSDRTFLVECL- 126
 QY 124 MEADKTVSAQIEVSEITTAFOQTIVMPV-----CRYEIDGGPTGQPVQFAIIQ 174
 Db 127 ---DNIT-----MPIVNRASFPADCTHLVMAEWSMSSEFQ-VGD 163
 QY 175 PVYHKWTC---DSETVDFPCAIVHSCFVDGNGDTVEILNADGALDKYLLNNLEFPTD 231
 Db 164 AIYHEWSCKLPNPAKTQTF---LTNCNALSONGCIHLIDENGCVIDSELMGDIVY---- 216
 QY 232 MAGCEAHVYKYADRSQLF-----YOCQISITIKENPNSC---VAPQCS 271
 Db 217 ---SDHVPLKYARARIFKFLTDKRIECLTEPCNNGSPCKDRVFPKCA 263

RESULT 30
 Q22724 PRELIMINARY; PRT; 711 AA.
 AC Q22724; 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-OCT-2001 (TREMBlrel. 18, last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, last annotation update)
 DE T24C2.1 protein.
 GN T24C2.1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]

RP SEQUENCE FROM N.A.
 RA Haynes C.;
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode C. elegans: A platform for
 investigating biology";
 RL Science 282:2012-2018(1998).
 DR EMBL; Z68120; CAA92199.2; .
 DR InterPro; IPR001507; Endoglin/CD105.
 DR SMART; SM00241; ZP; 1.
 DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; UNKNOWN_1.
 SQ SEQUENCE 711 AA; 80882 MW; 35AB81A46FE27E6 CRC64;

Query Match 10.0%; Score 202.5; DB 5; Length 711;
 Best Local Similarity 21.6%; Pred. No. 1.1e-10;
 Matches 63; Conservative 47; Mismatches 122; Indels 59; Gaps 10;

QY 5 LIAPCTLLIALSYIPVDNG-VESEPEIEGPGTSITINFTRNAFEHVVYKGLYDEGC 63
 Db 8 VIGIIFSVVLATSTDPDNAMHMGVPQVTCSAKLITVSFNTNIPFGCRISVFDKLTIPAC 67
 QY 64 RNDEGGROVAGISLPEDSCVARTSLNPRGIEFTTVVISFHLFTKYDRAVRVOCFY 123
 Db 68 NHDYSTNIQKNATFQMDILKCANPMLKNSRLRAYVEIGFHLVMTNSDRTFLVECL- 126
 QY 124 MEADKTVSAQIEVSEITTAFOQTIVMPV-----CRYEIDGGPTGQPVQFAIIQ 174
 Db 127 ---DNIT-----MPIVNRASFPADCTHLVMAEWSMSSEFQ-VGD 163
 QY 175 PVYHKWTC---DSETVDFPCAIVHSCFVDGNGDTVEILNADGALDKYLLNNLEFPTD 231
 Db 164 AIYHEWSCKLPNPAKTQTF---LTNCNALSONGCIHLIDENGCVIDSELMGDIVY---- 216
 QY 232 MAGCEAHVYKYADRSQLF-----YOCQISITIKENPNSC---VAPQCS 271
 Db 217 ---SDHVPLKYARARIFKFLTDKRIECLTEPCNNGSPCKDRVFPKCA 263

RESULT 31
 Q9W143 PRELIMINARY; PRT; 462 AA.
 AC Q9W143; 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, last annotation update)
 DE CG3541 protein.
 GN CG3541.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abail J.F., Adayant A., An H.-J., Andrews-Finnhock C., Baldwin D.,
 RA Balley R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokva D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
 RA Burlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dev I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glöck A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Mei M.-H., Ibegwam C.,
 RA Jalili M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Mostrefti A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pachl J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Relbert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spler E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svrtkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaverl J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of *Drosophila melanogaster*."
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003464; AAF47234.1; -
 DR FLYbase: FBgn0035030; CG3541.
 DR InterPro: IPR001507; EndoglIn/CD105.
 DR SMART: SM00241; ZP: 1.
 SQ SEQUENCE 462 AA; 50894 MW; 05F1AA173BEA558E CRC64;

Query Match 9.8%; Score 199; DB 5; Length 462;
 Best Local Similarity 22.2%; Pred. No. 1.3e-10;
 Matches 94; Conservative 55; Mismatches 150; Indels 124; Gaps 20;

QY 25 VGEDEIEEC--GPTSTTT-----NFTRNAFEGHYVVKLYOEGGRNDEGGROVAG--- 74
 DB 59 VPSVRIKICLSGSMLTITIDAPNHEH--GLFSGMITYPKGLSRNSCTLSEY--RDHGSLSR 115
 QY 75 ISLPDSCNVAFTSLNPRGIFVTYTVVISFHPLEVTYKDRAYRVQCFYMEADKTV---- 130
 DB 116 YLPLRSCN--TMRKETDGGIEFFNTIVYQPHLKLTIDGREGYHRCAKSKSDAAMKPKK 174
 QY 131 -----SAQIEVSEITTAFFQIYPM 150
 DB 175 YLRKHAQKQAFRRSDRREYGRSLDKQDDDEEDVDYANAPTOEEDVT---NNEIEM 230
 QY 151 PVC-----REIIDGGTGGPQVQFATIGQPVYHKMTCDSEYVDFCAVNSCEVDGN 203
 DB 231 PEGHMKITNDEHKIADVKIGDPLTIV-----SIDKQV--YGLHVTDCIVRDL 279
 QY 204 G-DTVEILNADGALDKYLLNLEYPITLMAGO---EAHVYKYADRSOLFYOCSITIK 259
 DB 280 GMEGRVLVGEDCPRDNEIMGOFNTQDRILANVTFPAHKFPYT--TSYTYCCNVRILCL 337
 QY 260 EPNSECVRPQCEPQGFQAVKTGGAAPRAAAQDLRLKRSAPENIT--IDVETDINT 316
 DB 338 EDPPTQEARPQCS-----GKRPKROAAD-----SKEDGLPATIEVSGLYV 379
 QY 317 LEISDQNALPYDLRHRALLQHNQGPVILAAVQNGICSPFGFSFMG-----LSIALIA 371
 DB 380 NENENAN-----DSDEDAVYKER-----TTDDALCVSQTFTATAIATIGIILMAVYA 427

RESULT 32
 Q9GCUH4 PRELIMINARY; PRT; 284 AA.
 AC Q9GCUH4
 DT 01-MAR-2001 (TReMBLrel. 16, Created)

DT 01-MAR-2001 (TReMBLrel. 16, last sequence update)
 DT 01-MAR-2002 (TReMBLrel. 20, last annotation update)
 DE Hypothetical protein Y55D5A.a (Y55D5A.6 protein).
 GN Y55D5A.A OR Y55D5A.6.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
 OC Rhabditidae; Felodierinae; Caenorhabditis.
 OX NCBI_TaxId=6239;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RX MEDLINE=99069613; PubMed=9851916;
 RA None;
 RT "Genome sequence of the nematode *C. elegans*: a platform for
 RT investigating biology. The *C. elegans* Sequencing Consortium."
 RL Science 282:2012-2018(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RA Waterston R.;
 RL Submitted (Oct-2000) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RA Waterston R.;
 RL Submitted (Apr-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AC084196; AAG23450.1; -
 DR EMBL: AC084196; AAK29949.2; -
 DR InterPro: IPR001507; EndoglIn/CD105.
 DR SMART: SM00241; ZP: 1.
 SQ SEQUENCE 284 AA; 32361 MW; EC459689E4661283 CRC64;

Query Match 8.6%; Score 174; DB 5; Length 284;
 Best Local Similarity 23.9%; Pred. No. 1.8e-08;
 Matches 73; Conservative 50; Mismatches 126; Indels 56; Gaps 15;

QY 101 VVISFHPLEVTYKDRAYRVQCFYMEADKTVSAQIE--VSE---ITTAFTQTIYMPVCRY 155
 DB 8 VIVAFHPPLTVTSDFRAHCFEDEFKRTKEIGIELNLQEHLLGNFQ----LPKISM 62
 QY 156 EILDGP---TQCPQFAL-----IGQPVYHKMTCDSEYVDTTCVAVNSCFVDDGND 205
 DB 63 HILPAGESLTTKTQNFANEQKVLNVGDPIMEFKLEQEH--GIQGIOLERSASSENK 121
 QY 206 TVEILNADGALDKYLLNLEYPITLMAGOEAHV--KYADRSOLFYOCSITIKE--PNS 263
 DB 122 GKKIIE--NCSIDBELISDTTYSQPSKIYATSLAFKPEBEIVTRCAVRCVKKTEHL 180
 QY 264 ECVRPQ---CSEPQGFQAVKTGGAAPRAAAQDLRLKRSAPENIIDVETDINTLEI 319
 DB 181 EILNGEEDLCSANNCGFVSPRESSOR-----SRQLVGKNNNNTRTDI----- 224
 QY 320 SPDQNALPYDLRHRALLQHNQGPVILAAVQNGICSPFGFSFMGLSTALIAVY---IT 376
 DB 225 -----IYVNGRFVREKQSSPQEA--RATSTQTEFCMPDIVY--YGLSTVVICYLTISTT 275
 QY 377 ISFKF 381
 DB 276 VVFKF 280

RESULT 33
 Q9VZC5
 ID Q9VZC5 PRELIMINARY; PRT; 604 AA.
 AC Q9VZC5.
 DT 01-MAY-2000 (TReMBLrel. 13, Created)
 DT 01-MAY-2000 (TReMBLrel. 13, last sequence update)
 DT 01-JUN-2002 (TReMBLrel. 21, last annotation update)
 DE CG15020 protein.
 GN CG15020.
 OS *Drosophila melanogaster* (fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Db 449 TGSPALYTECVRM-----CH-----GRCPSQPCRMRLKAVTKR-- 483

QY 303 EPENIIDVRTDINTLEISDDNQLPVD 329
: : : : :
Db 484 DTSMN--TATNISITPLSADGELTTE 508

RESULT 34

Q9BMD4 PRELIMINARY; PRT; 1549 AA.

AC Q9BMD4
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE No-mechanoreceptor potential A short isoform precursor.
GN NOMPRA OR CG13207.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Phrygotea; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
CC Empyridae; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21134724; PubMed=11239432;
RA Chung Y.D., Zhu J., Han Y., Kernan M.J.;
RT "nompA encodes a PNS-specific, 2P domain protein required to connect mechanosensory dendrites to sensory structures.";
RL Neuron 29:415-428(2001).
DR EMBL: AF334032; AKR09434.1; -
DR FlyBase: FBgn0016047; nompa.
DR InterPro: IPR001064; Crystallin.
DR InterPro: IPR001507; EndoglIn/CDI05.
DR InterPro: IPR003014; PAN.
DR InterPro: IPR003609; Pan_app.
DR Pfam: PF00024; PAN_5.
DR SMART: SMO0241; PAN_AP_4.
DR SMART: SMO0241; ZP_1.
DR PROSITE: PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
KW Receptor; Signal.
FT SIGNAL 1 21 POTENTIAL.
SQ SEQUENCE 1549 AA; 174271 MW; 78631BCACA39D5B93 CRC64;

Query Match 8.0%; Score 163; DB 5; Length 1549;
Best Local Similarity 19.8%; Pred. No. 2e-06;
Matches 85; Conservative 62; Mismatches 155; Indels 128; Gaps 16;

QY 27 GEPEIECGPTSTTN-----FNTRNMFEGHYVKKGLDYDGECNDGGRGVAGISLP- 78
: : : : :
Db 1013 GRSDSECDVOTCEEGMEETTIRTBEGFLRIYYGYDFRCFFNGGVNVLRSLSGPQ 1072

QY 79 -FDSCGNVARTSRSLNRGIFFVTTVVISPHLFVTKVDARVROCP----- 122
: : : : :
Db 1073 GYPDCGTOR-----YGTLTINIYVVGSDNVQTSRDKRNYLTCTFRPGEAYVSSGIIG 1126

QY 123 -----YMADKTIVSAOIEVSEITTAFOQTQIVPMVCRYEILTDGPT----GOVP 167
: : : : :
Db 1127 AGSGSPRIETELPAENTLSKVRLS-----ILYGRPTTTIAVGDPL 1168

QY 168 QFAIIIGOVYVHKWTCDSFTVFPCAVVHSCFVDGDNQDTVELIANDGALDKYLNNLEY 227
: : : : :
Db 1169 TFRLEAQOGYNHVT-----DIFATNVVA--RDPSYSGRSIQIDRFEGCPVDPFPPELDK 1220

QY 228 PTD---LMAAGEAHYKYKAADSOLFYOCAISITIKEPSSECVRPCSPPOGGAVKTGCA 284
: : : : :
Db 1221 LRDGDTLEA--RFNAFKIPESNFVFEA-----TVNSCEGQPAPCBPAG----- 1265

QY 285 AKPAAAAOLRLLKRS-AEPD-----NIIDVRDINTLEISDDNQLPVD-- 329
: : : : :
Db 1266 ROEPFGRRRSLSNTTELPEPALALESSQLAEASTLDDEVVNSTVSATLGQVLNMT 1325

QY 330 -----LRHALLOHMGQPIITAAVONGICMSPFPGISMFGJLSIA 368
: : : : :
Db 1326 QLGEKTKTEEBEQVREMIYEFTTEELESKEYSRKLYAAPVETVCMTP---AEYGLITTA 1382

```
QY      369 LIAAVITIS 378
          :|:|:|
Db      1383 IILLMILFS 1392
```

RESULT	35			
Q9BMD5				
ID	Q9BMD5	PRELIMINARY:	PRT:	1557 AA.
AC	Q9BMD5;			
DT	01-JUN-2001 (TrEMBLrel. 17, Created)			
DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)			
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)			
DE	No-mechanoreceptor potential A long isoform precursor.			
GN	NOMPA OR CG13207.			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha			
OC	Ephydroidea; Drosophilidae; Drosophila.			
OX	NCBI_TaxID=7227;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=1134724; PubMed=11239432;			
RA	Chung Y.D., Zhu J., Han Y., Kernan M.J.;			
RT	"nompa encodes a PNS-specific, zp domain protein required to connect			
RT	mechanosensory dendrites to sensory structures.";			
RL	Neuron 29:415-428(2001).			
RL	EMBL: AF34031; AAK09433.1; "			
DR	FLYBase: FBgn0016047; nompa.			
DR	InterPro: IPR001064; Crystallin.			
DR	InterPro: IPR001507; Endoglin/CD105.			
DR	InterPro: IPR003014; PAN.			
DR	InterPro: IPR003609; Pan_app.			
DR	Pfam: PF00024; PAN; 5.			
DR	SMART: SM00473; PAN_AP; 4.			
DR	SMART: PS00241; zp; 1.			
DR	PROSITE: PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.			
DR	Receptor; Signal.			
ST	SIGNAL	1		
ST		21	POTENTIAL.	
Q1	SEQUENCE	1557 AA;	175291 MW;	134BDBBA89245183 CRC64;

QY	369	LIAVITIS	378
	:	↑	:
Db	1383	IILLMILFS	1392

RESULT 36		
09V533	PRELIMINARY;	PRT; 1638 AA.
ID	09V533	
AC	09V533;	
DT	01-MAY-2000 (TREMBLrel. 13, Created)	
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)	
DT	01-MAY-2002 (TREMBLrel. 20, Last annotation update)	
DE	CG13207 protein.	
GN	NOMPA OR CG13207.	
OS	Drosophila melanogaster (Fruit fly).	
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;	
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;	
OC	Ephyridioidea; Drosophilidae; Drosophila.	
OX	NCBI_TaxID=7227;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	SPRAIN-BERKELEY.	
RX	MEDLINE=20196006; PubMed=10731132;	
RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,	
RA	Amantides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,	
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,	
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,	
RA	Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,	
RA	Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos D.L.G.,	
RA	Abbil J.F., Aghayani A., An H.-T., Andrews-Plambeck C., Baldwin D.,	
RA	Balcer R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,	
RA	Besson K.Y., Benos P.V., Bouck J., Brokstein P., Brothier P.,	
RA	Botkova D., Botchan M.R., Beum J., Brokstein P., Brothier P.,	
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,	
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,	
RA	de Pallos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,	
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,	
RA	Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,	
RA	Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,	
RA	Gloder C., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,	
RA	Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,	
RA	Hoslin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwu C.,	
RA	Jajalil M., Kalish F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A.,	
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,	
RA	Lasko P., Lai Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,	
RA	Lin X., Matei B., McIntosh T.C., McLeod M.P., McPherson D.,	
RA	Merkulov G., Mlshina N.V., Modyarty C., Morris J., Moshrefi A.,	
RA	Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,	
RA	Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,	
RA	Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,	
RA	Reinelt K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,	
RA	Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,	
RA	Splet E., Spradling A.C., Stapleton M., Strong R., Sun E.,	
RA	Styrskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,	
RA	Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weisenbach J.,	
RA	Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,	
RA	Ye X., Yeh R.-F., Zaveri J., Zhou M., Zhu M., Zhu G., Zhao Q., Zheng L.,	
RA	Zheng X.H., Zhong F., Zhong W., Zhou X., Zhang S., Zhang L.,	
RA	Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,	
RT	"The genome sequence of Drosophila melanogaster."	
RL	Science 287:2185-2195(2000).	
DR	EMBL: AEO03826; AAF5668.1; -	
DR	FLYBase: FBgn0016047; nompa.	
DR	InterPro: IPR001064; Crystallin.	
DR	InterPro: IPR001507; Endoglin/CD105.	
DR	InterPro: IPR003014; PAN.	
DR	InterPro: IPR003609; Pan_app.	
DR	Pfam: PF00024; PAN_5	
DR	SMART: SM00473; PAN_AP; 4.	
DR	SMART: SM00241; ZP_1	
DR	PROSITE: PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.	
DR	SEQUENCE 1638 AA; 184367 MW; 991620EABFA3EB15 CRC64;	

[illegible]

RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
 RA Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Watson L., Wilkinson-Spratt J., Wohlman P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.";
 RL Nature 368:32-38(1994).
 RN
 RP
 RC SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Scheet P.;
 RT "The sequence of C. elegans cosmid T21H3.";
 RL Submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.
 RN
 RP
 RC SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Waterston R.;
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF016429; AAB65365.1;
 DR InterPro: IPR001507; EndoglIn/CD105.
 DR SMART: SM00241; 2P; 1.
 SO SEQUENCE 514 AA; 59123 MW; 1F9E9BDE710D991A CRC64;

Query Match 7.5%; Score 151.5; DB 5; Length 514;
 Best Local Similarity 19.0%; Pred. No. 6e-06;
 Matches 67; Conservative 63; Mismatches 108; Indels 115; Gaps 16;

QY 82 CAVARFRSLNPRGIFVTYV--VISFHPFLFVKVDRAYVOC-FYWEADKTVSAQIEVSE 138
 DB 113 CQVRRAYKQVPS-FVTYSIRLVSHNSDELTEFDKIYVOTSLHL--KTMAYETSSPO 168
 QY 139 ITT---AFOTQIVPWCYREYILD---GSP-----TGSP----- 166
 DB 169 TTTTSSVTHSNAGRPCKKYSLHTRVGGPRPVSAHGEVKKATKIQMKSLALHM 228
 QY 167 -----VOFAITGO----- 181
 DB 229 VKNELPNSGNFFVKKYPRHDSLCYMRFSILNYSKEYGSRNSKNTINDISMHNFOVYHRMK 288
 QY 182 C-----DSEYDTPCAVYHSCFVDDGNDGTVIELNADSCALDKYLLNLEYPDTLMAG- 234
 DB 289 CAAPLYENRQMPMLIFKYACVYHDENRRTAIIIDNCGSLDEITIPPEY-DIENGV 346
 QY 235 --QEAHVYKADRSOLFYOQIISI-TIKEPN-SECVRPOCSPEGFGAVKGTGAAPAA 290
 DB 347 IYTPSKAFEPFANSNHVFCMISVCSAVIDPSCRSNVPKPC-----KNGLAKKKRQ 397
 QY 291 AAGLLKRRSAPEPNIIDVTRDINTLEISDNOALPVLRHALLQHNQOPY 343
 DB 398 LPEEMTIEQRLRLRIHELMKVKRENTIVEGRPT-----KLLSHGAVPL 441

RESULT 39
 QYV9X1 PRELIMINARY; PRT; 758 AA.
 AC QYV9X1;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, last annotation update)
 DE CG12063 protein.
 GN
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Anthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OK NCBI_TaxID=7227;
 RN
 RP
 RC SEQUENCE FROM N.A.
 RC STRAIN-BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Cealniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Branton R.C., Rogers Y.-H.C., Blazel R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abriil J.F., Agdayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktarovu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brocksiepen P., Brottler P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Honck J.,
 RA Hostin D., Houston K.A., Howland T.J., Mei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lascko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Morkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,
 RA Palazzolo M., Peltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Relaeert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kimios I., Simpson M., Skupski M.P., Smith T.,
 RA Splier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissenhach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 DR EMBL; AE003778; AAF57158.1;
 DR Flybase; FBgn0039851; CG12063.
 DR InterPro: IPR001507; EndoglIn/CD105.
 DR InterPro: IPR003014; PAN.
 DR InterPro: IPR003609; Pan_app.
 DR Pfam; PF00024; PAN; 3.
 DR SMART; SM00473; PAN_AP; 3.
 DR SMART; SM00241; 2P; 1.
 SO SEQUENCE 758 AA; 81720 MW; A7823019671861E3 CRC64;

Query Match 7.4%; Score 150; DB 5; Length 758;
 Best Local Similarity 23.6%; Pred. No. 1.4e-05;
 Matches 81; Conservative 50; Mismatches 114; Indels 98; Gaps 20;

QY 31 IECGPTSTITINFTNFAFGHYVYKGLYDQECRNDGGRQYAGISLPD-----SC 82
 DB 393 IDCRSGEMTKITRTSKLFGKYAKGA--PKSCAVVNN-----SLFEDLMRYNDLEC 444
 QY 83 NVART---KSLNPRGIFVTYVVISFHPFLFVKVDRAYVOCFYMAADKTVSAQIEV--- 136
 DB 445 NVRSQSAVGRYXN-----DIVIQHDMIVYSSDGLAVSQCYDITNTVANNNDLGVF 496
 QY 137 SETTTFQTOIY---PMPYCRYEIIDGGPTGQVPVQFAIGQYVYHKWTCDSFTVPT---F 190
 DB 497 GEIESTLSEIIVDSPNVIMKTITARDG---SDMKRAIEVGPLALRF-----EIVDANSY 549
 QY 191 CAVHSCFVDDGNDGTVIEL-LNADSCALDKYLLNLEYPDTLMAGQEAHVYKKA-DRSQ 247
 DB 550 EIFVRELVAIDGFI-DSAEITLLDANCPCPDXYIMSAMQ-----KLANNRKV 594
 QY 248 LFYQCOISITITEPNSECR-----PQSEP-----QGGAAYK----- 280
 DB 595 LLSQFD---AFKFPSELVQFALVTPCIPRC-EPYICDNDENGELKSLSYGRRKRSYL 650
 QY 281 --TGGAAPKPAALALRLKRSAPENIIDVTRDINTLEISD 321
 DB 651 NGTDGEVLAIKSERQKRDVSHQAAGDENILVQ-----SIQITD 689

THIS PAGE BLANK (USPTO)